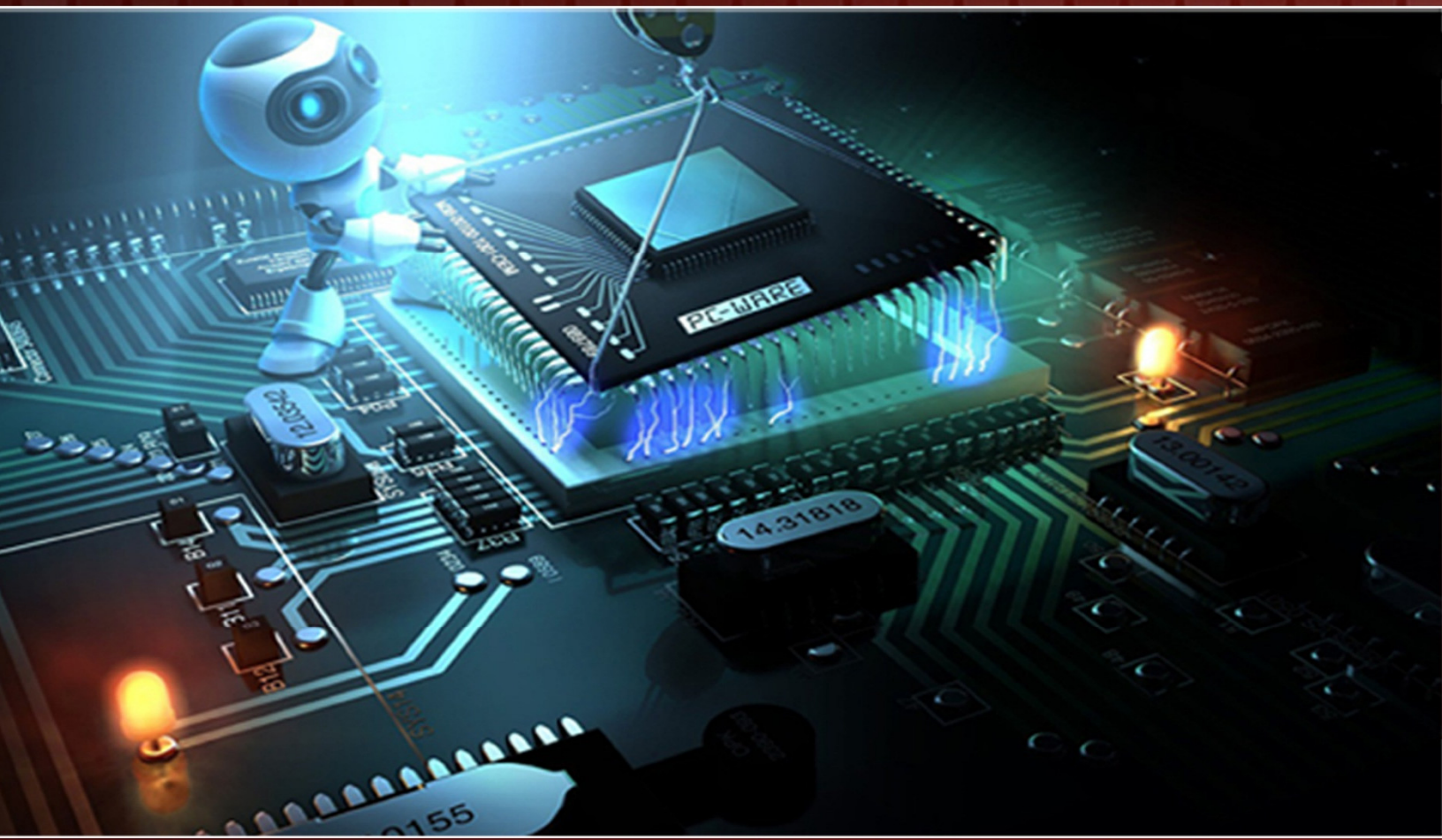
ISSN: 2054-7390

SPECIAL ISSUE ON

INTERNATIONAL CONFERENCE ON AFFECTIVE COMPUTING, MACHINE LEARNING AND INTELLIGENT SYSTEMS (ACMLIS)

TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE



Society for Science and Education
UNITED KINGDOM

EDITORIAL ADVISORY BOARD

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone's views or feelings etc.

TABLE OF CONTENTS

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Resources Indexation-Based E-Learning System

[1]Ibtissam Azzi, [2]Abdelhay Radouane, [3]Adil Jeghal, [4]Hamid Tairi
[1]Laboratoire LIIAN, Faculté des Sciences Dhar Mehraz, U.S.M.B.A. Fès, Maroc
[2]Département d'informatique, Centre Régional des métiers de l'éducation et de Formation, C.R.M.E.F. Fès, Maroc, 3Laboratoire LIIAN, Faculté des Sciences Dhar Mehraz, U.S.M.B.A. Fès, Maroc
[4]Laboratoire LIIAN, Faculté des Sciences Dhar Mehraz, U.S.M.B.A. Fès, Maroc
[1]Azzi.ibtissam@usmba.ac.ma, [2]jeghal.adil@usmba.ac.ma, [4]Htairi@yahoo.fr

ABSTRACT

We propose in this article a study that illustrates the techniques used to represent the educational objects which serve to facilitate their reuse. We will enrich this study by taking into account the semantics of the contents of the learning objects while including the metadata as parameters of indexing. Another contribution of our study concerns the fact that this indexation relies on ontologies which allow a better semantic representation and facilitate communication between the machine and the users. On the other hand, we will discuss the current standards of indexing while discussing the cases of their use to try to extend the use of these standards in other situations requiring the indexing of resources in E -Learning systems.

Keywords- Resources indexation; Metadata; Ontology; Standards; E Learning.

1 Introduction

The deployment of information and communication technologies has taken a major boom in the professional and personal domains of individuals, and computing has become a necessary and powerful medium for improvement and innovation. Indeed, with the era of the Internet, the field of education and training has found a room to optimization with e-learning which relies on the provision of learning content through learning designs in a digital environment. However, with the exponential growth in the number of available resources, it is increasingly difficult for learners and teachers to find the learning content that best meets their needs. To enable localization, access, sharing and re-use of these resources, it is essential to describe them. One solution may come from the use of information derived from the descriptions of learning resources. On the other hand, besides content development, the creation of such resources poses many difficulties for its designers. The use of these resources in learning designs implies not only that the resource be divided into coherent learning objects but also the activities associated with the pedagogical objectives be defined. Existing resources are difficult to re-use because they have not been designed for this purpose; they often correspond to simple on-line presentations of documents that have not been created specifically for their use in learning environments. [1] finds insufficient or no application of a pedagogical approach, whether in the presentation of learning resources or the sequencing of learning activities in current tools. On the other hand, to be reusable, a resource must be indexed so as to know at the same time the content, the educational objectives aimed at, the conditions of reusability, etc. In the literature one finds standards that offer solutions for the problem of indexing in the framework

DOI: 10.14738/tmlai.54.2967
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2967

of E- Learning. We will discuss their advantages and limitations in this article which is structured in the following way: in section II we will clarify the value of introducing resource indexing in the solution of problems related mainly to the personalization of learning in systems of E -Learning. Subsequently, in Section III, we will discuss the role of metadata in the semantic description of the resources while giving an overview on the different standards using the metadata. We will also discuss in section IV the notion of ontologies where we will discuss the various difficulties encountered in their conception. Section V describes the reuse of the ontologies. On the contrary in the conclusion, we will present our point of view concerning the exploitation of ontologies in the indexation of resources in particular, we will exploit the important property of ontologies namely the property of integration which allows the development of ontologies already designed to create another. We want to create an ontology that takes into account the learner's learning style in order to personalize it. This will be a promising prospect in our future work.

2 Resource Indexation In E Learning Systems

The definition of a learning object has been the subject of several debates. According to the Learning Object Metadata (LOM) standard [2], a learning object is defined as any numeric or non-numerical entity that can be used, reused or referenced during learning activities. This definition is considered by [3] to be very broad because it may include an object, a person or an idea. [4] refined this definition by adding that a learning object is a unit of learning content that is independent, autonomous and predisposed to be reused in multiple pedagogical contexts. [5] emphasizes that the notion of learning object remains unclear, because some definitions are either too broad or too narrow in relation to the use that can be made of them. In order to remedy this problem of definition we consider that a learning object is a material that can be selected, combined with another material according to the needs of the teachers and the learners. It is also a learning content which must exist as such, and which can be easily searched and indexed [6]. To enable indexing and retrieval of these learning objects, it is necessary to use a set of metadata. The fundamental idea behind the creation of learning objects is the possibility of building even small components or elements that can be reused several times in different learning contexts. [7] argues that, often when teachers or creators of learning content first access learning materials, they decompose it into its constituent elements and then assemble them in order to construct one which supports their pedagogical objectives. If teachers or content creators could have these resources as small units, this first stage of decomposition would be bypassed. This can greatly increase the speed and effectiveness of creating new resources [3]. However, in order for these small units to be reused they must be able to find them. In order to facilitate the adoption of the "learning object" approach, the IEEE 2 consortium created the Learning Technology Standards Committee (LTSC) in 1996 to develop and promote standards for learning technology. The committee then chose the term learning object to describe these small reusable units. The indexing of a resource makes it possible to describe it for better use because a non-indexed resource is an unexploited resource which is impossible to find.

3 Metadata And Lom, Scorm Standards

Metadata is semantic data that describes resources. Tim Berners Lee, the inventor of the web, said that digital information resources must be understandable by human beings and processable by software. The indexation must then be precise, which sometimes implies a very large number of fields to fill, more than fifty in the LOM. Some of the indexing can be done automatically, such as automatic recognition of the

file format by the software. But the description of the pedagogical objective of a resource, for example, should be done manually.

Next, the display of the metadata fields will be adapted according to the type of user: a user will not need to see the same fields as an indexer. On the web, search engines like Google in principle make it possible to find all that is accessible. This shows that metadata is really needed. They help not only to find resources more easily, but also to evaluate them, even if they are not the same descriptors that are used in both cases. Finally, metadata also facilitates maintenance, through the management of different versions of a document, and automatic assembly of resources. It is then essential to anticipate and think about all the uses of a resource before defining the set of associated metadata. On the other hand, indexing is complex and requires some standardization. Indeed, in certain contexts, it is possible to impose the indexing system, the descriptors, the software, the file formats, the operating system, the instantiation of the Learning Object Metadata (LOM) standard. However, a change in context will then encounter some problems and involve transforming data. In other cases, the problem is rather how to interact with other platforms. The aim is therefore not to repeat the indexing each time a change of platform or interlocutors occurs. The use of standards or standards is then made mandatory. Indeed, standards are the only means that allow interoperability and evolution of systems over time. A norm or a standard can not necessarily meet everyone's needs, but must be scalable. Norms and standards must be built on a basis of common needs. Similarly, for indexing to survive software and hardware, it must imperatively be based on standards. Transcripts allowing the passage from one environment to another will necessarily be written when norms or standards evolve. Standards and norms therefore make it possible to maintain resources and, thanks to indexation, they can be guaranteed accessibility and interoperability through exchanges between systems. Resources can then be reused and adapted.

The current e-Learning standards offer a partial solution for reusability. In particular, SCORM (Sharable Content Object Reference Model) [8] and LOM (Learning Object Metadata) help to homogenize representations of learning resources and facilitate interoperability. LOM proposes different metadata necessary for the description of the learning resources, but this description is not enough, it lacks the semantic representation of the contents. SCORM for its part, allows the structuring of the contents of the learning resources and their relations with the environment of use. However, the representations obtained via these standards are not sufficient to allow and ensure the reusability of resources or parts of resources because they take into account neither the semantic content of the metadata they introduce nor the relations that exist between each learning object. To remedy these problems, representations of "learning objects" which are based on a representation of knowledge in the form of ontologies have been proposed. They deal with the reusability aspect of these documents, in particular their semantic indexing, the pedagogical aspect which makes it possible to learn the knowledge conveyed in the documents.

3.1 LOM (Learning Object Metadata)

At present, the most complete and most suitable standard for the field of education and e-learning remains the LOM.

The LOM or Metadata for Educational Objects standard is a standard developed in 2002 by the IEEE consortium, which defines the structure of a metadata instance for an educational object. It consists of nine categories (general, life cycle, meta-metadata, educational, technical, rights, relation, annotation, classification) each performing a different function. The descriptors of the LOM can be used in the design

of E-learning systems for indexing learning objects. For this they must be implemented in a structured language. The representation of the abstract model in a specific format is called "binding".

For LOM metadata, there are two metadata: XML binding and RDF binding. The use of a language such as XML (Extensible Markup Language) makes it possible to perceive the structure of the LOM and facilitates the exchange of metadata between different systems. If the XML binding of the LOM has the merit of being easy to implement, it remains insufficient for the representation of the elements of the LOM since it does not make it possible to express the semantics of these elements.

While XML structures metadata, RDF (Resource Description Framework), poses on XML a framework for defining these metadata. This is why RDF is referred to as metalanguage metadata. It is intended to structure the information accessible on the web and to index it effectively. However, whether they are implemented in XML or RDF, using LOM metadata poses a number of problems. Indeed, according to [9] to manage resources with the metadata one can enumerate many problems encountered when using the metadata of the LOM (complexity of the structure, important number of elements, difficulty to inform some fields, etc.).

3.2 SCORM (Sharable Content Object Reference Model)

It is a model for the assembly of web contents and a learning environment for learning objects. Its aim is to put in place the right structure of the content of the course and its interactions with its environment. The structuring of the content of the teaching modules according to the SCORM model makes it possible to reuse them in other modules for different training or systems. Moreover, it improves the dialogue between the learning objects and the system on the one hand, and between the actors and the system on the other.

SCORM deals with several elements, for example:

- Packaging: its objective is the transmission of content from one platform to another, the import or export of contents learning objects to make them available to others. It is equally interested in the structuring of learning objects.
- Meta data: they are from LOM and aim to share standard information that describes the nature and purpose of the content.
- Communication or execution environment: determines the communication with a web environment.
- Sequencing and navigation: defines a method of representing navigation between learning objects.

4 Ontologies

In order to solve some of the problems associated with lack of semantics, ontologies have been introduced [10], [11], [12] and [13]. According to [14], an ontology groups the concepts that represent the whole knowledge of a domain into an explicit and formal specification. It shows the relations as well as the association rules that exist between these concepts in order to allow, on the one hand, the computer the production of new knowledge by means of an inference and, on the other hand, to allow the man and the computer to give common meanings to the terms used in a field of activity in order to remove any ambiguity during the treatments.

An ontology is translated by the following elements:

- Concepts: also called terms or classes of ontology, correspond to the relevant abstractions of a segment of reality (the domain of the problem), selected according to the objectives that are set and the application envisaged for the " ontology.
- Relationships: A relationship is a link between concepts and describes a type of interaction between concepts.
- The axioms: constitute assertions, accepted as true, about the abstractions of the domain translated by the ontology.
- The instances: constitute the extensional definition of ontology; these objects convey knowledge (static, factual) about the domain of the problem.

However, the design of an ontology is a matter of ontological engineering [15]. In the literature, we can find several methodologies for designing ontologies dedicated to E -Learning [16], [17], [18], [19] and [20]. Reference [21] proposes an ontology of courses that breaks down into 3 sub-parts describing the content, context and structure. The content is commonly referred to as the ontology of the domain. The other two parts are related to the pedagogical aspect (structuring in chapters, nature of the parts, etc.). Reference [22] considers that domain and pedagogy can be combined within a single ontology. The ontology provides a set of Concepts and relationships modeling a domain, a pedagogical strategy, etc. Therefore this requires a modeling language. One can cite the UML (Unified Modeling Language). One of the advantages of the UML schema compared to other formalisms is that it offers a direct connection to the world of software engineering. In addition, there are already tools to manipulate the models. From these diagrams are derived a vocabulary and a metadata schema for indexing resources in XML with respect to this model The other favours which ontologies offer lies in their property of integration. Indeed, one can use ontologies already conceived to conceive a new ontology adapted to our needs.

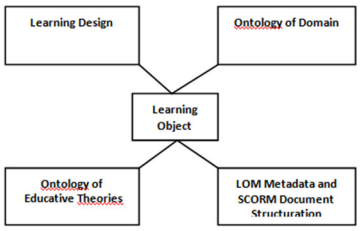
5 Reuse of ontologies

In this context, work highlighted the possibility of re-use of certain ontologies. Ontologies of field and task, nonspecific to a given application, can be regarded as reusable [23], in particular if they specialize the concepts of a high level ontology. In [24], we identified 4 stages which the originator must carry out to use of existing ontologies:

- Specification and formalization of the needs.
- Specification and the formalization of the requirements in terms for operational features for the software.
- Appropriation of ontologies and creation of an ontology of application.
- It is a question of re-using an existing ontology of field or an existing ontology of task and then to create an ontology of application by using these two ontologies.
- Operationnalisation.
- The operationnalisation of ontology thus consists of the computer specification of the operations applicable to the concepts in an operational language.
- Finalization of the establishment and data model.

It is a question of choosing the environment of establishment of the application, with regard to the management of this data and their establishment then the originator/developer establishes his interfaces and its application in a language targets (e.g.:Java).

It is in this direction that we wish to act in our future work while trying to use ontologies of field and ontologies of pedagogy to create an ontology which gives an account of the preferences of the user in particular his learning style. For that, we will adopt the model which represents knowledge relative to the various aspects of a document and its uses. It is described by the following schema [25] and [26]:



6 Conclusion

In this article, we have clarified the value of indexing resources in E -Learning systems and have also given an overview of the problem of indexing in these systems. In addition, we have overflowed the various descriptions existing in the literature capable of indexing the resources. In this sense, we discussed the metadata and their ability to describe the semantics of resources in particular, we have dealt with the LOM and SCORM standards where we have recalled the advantages and limitations of these standards. We have also clarified the need to introduce ontologies as a means of remedying the deficiencies encountered in the description made by the metadata. Despite these different constraints, and because the scope of ontologies is wide, and the problems related to e-learning are diversified, we wish to develop an ontology in our future work, taking into account the learner's learning style. This ontology will be associated with LOM metadata.

REFERENCES

[1] Psyché V., Bourdeau J., Nkambou R. et Mizoguchi R.”Making Learning Design Standards Work with an Ontology of Educational Theories”. Artificial. Intelligence in Education (AIED 2005), 725-731.

[2] LOM (2002). LOM standard, document IEEE 1484.12.1-2002.

[3] WILEY D.” Connecting learning objects to instructional design theory : A definition, a metaphor, and a taxonomy”. The Instructional Use of Learning Objects, p. 571–577, 2002.

[4] POLSANI P.”Use and Abuse of Reusable Learning Objects”. Journal of Digital Information, 3(4), 2003–02.

[5] BOURDA Y. ’’Objets pedagogiques, vous avez dit objets pedagogiques ?’’ Cahier GUTenberg n 39, 40, 71–79, 2001.

[6] ABEL M.-H. ’’Apport des Memoires Organisationnelles dans un contexte d’apprentissage’’. Habilitation a diriger des recherches, UTC , 2007.

[7] REIGELUTH C. ’’A New Paradigm of ISD ? ’’Educational Technology –SADDLE BROOK NJ-, 36, 13–20, 1996.

[8] SCORM. Le modèle SCORM (2004). http://www.adlnet.org

[9] MICHEL A. ''La gestion des ressources avec les métadonnées''. Journee Normes et standards educatifs, 26, 2004.

[10] Gruber T.'' A transaction approach to portable ontology specification '', Knowledge Acquisition, 5(2) p. 199-220. 1993

[11] Uschold M., Gruninger M.'' ONTOLOGIES : Principles, Methodes and Applications'', Knowledge Engineering Review. vol. 11, n° 2, 1996.

[12] Borst W.N.''Construction of Engineering Ontologies'', PhD Thesis, University of Twente, Enschede, 1997.

[13] Giraldo G., Reynaud C.'' Construction semi-automatique d'ontologies à partir de DTDs relatives à un même domaine'', Actes de IC'2002, Rouen, mai 2002.

[14] Studer R., Benjamins VR. et Fensel D. ''Knowledge Engineering: Principles and Methods''. Data and Knowledge Engineering (DKE), 25(1-2), p 161-197, 1998.

[15] GÓMEZ-PÉREZ A. ''Ontological Engineering''. Tutorial ECAI'98, Budapest, Hongrie, août 1998.

[16] Desmoulins C. et Grandbastien M. ''Des ontologies pour la conception de manuels de formation à partir de documents techniques''. Revue Sciences et Techniques éducatives, 2002.

[17] Duitama J. F., Defude B., Bouzeghoub A., Carpentier C. ''A Framework for the generation of Adaptive Courses Based on Semantic Metadata'', Proc. MAWIS Workshop, with OOIS'03 Conference, Geneva, September, 2003.

[18] Dehors S., Faron Zucker C., Giboin A., Stromboni J. P. ''Un Web Sémantique de Formation par Questionnement''. Acte des journnées Extraction et Gestion des Connaissances 2005. EGC 2005

[19] Ranwez S. et Crampes M. ''Conceptual Documents and Hypertext Documents are two Different Forms of Virtual Document'', Workshop "on Virtual Documents Hypertext Funcionality and the Web" of the 8th World Wide Web Conf., Toronto, 1999.

[20] Iksal S. ''Spécification Déclarative et Composition Sémantique pour des Documents Virtuels Personnalisables'', Thèse de Doctorat, EHESS, Brest, France, 2002.

[21] STOJANOVIC L., STAAB S. & STUDER R. ''Elearning based on the semantic web'', WebNet2001 - World Conference on the WWW and Internet, Orlando, Florida, USA, 2001-10.

[22] ABEL M-H. ET AL. '' Using an Organizational Memory for e-learning'', Workshop of Knowledge Management and Organizational Memories, ECAI'2004, Valencia, Spain, 2004- 08-22.

[23] Benayache, A.''Construction d'une mémoire organisationnelle de formation et évaluation dans un contexte e-learning''. le projet MEMORAe. Thèse de doctorat de l'université de technologie de Compiègne, 2005.

[24] Pascal Lando.''Conception et développement d'applications informatiques utilisant des ontologies'': application aux EIAH, rjc_eiah, Mai, 2006.

[25] Nathalie Hernandez, Josiane Mothe, Bachelin Ralalason, Patricia tolf, CIDE.9.''Modèle de représentation sémantique des documents électroniques pour leur réutilisabilité dans l'apprentissage en ligne''. Fribourg, Suisse, 18-20 September 2006.

[26] Nathalie Hernandez, Josiane Mothe, Andriantiana Bertin Olivier Ramamonjisoa, Bachelin Ralalason, Patricia Stolf. ''Indexation multi-facettes des ressources pédagogiques pour faciliter leur ré-utilisation'', vol. RNTI-E-12, pp.207-230, 2008.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Data Analysis Application to Investigate Relationships between Metacognition and Learning Styles

Anass Hsissi, Hakim Allali, Abdelmajid Hajami
LAVETE – e-learning Research Lab, University of science & technology, University Hassan I, SETTAT, Morocco
anasshsissi@gmail.com

ABSTRACT

The purpose of this study is to examine if learning styles predict effectiveness in learning. Participants of the study consisted of 80 students selected from different classes of the university of science and Technology –Settat, Morocco, two questionnaires on learning styles and metacognition related to reading and comprehension were handed out to members of the sample during academic year 2016/2017. Data analysis techniques were used to understand our dataset; principal component analysis was used to discover significant patterns within data. Results indicate that there are significant relations between the dimensions of learning styles, metacognition and performance in comprehension activity. Results also indicate that learning styles significantly explain and predict all sub-dimensions of metacognition.

As perspective for our research, we tend to offer to each student the appropriate environment and conditions to enhance learning, in the conclusion we will discuss the implications and recommendations for designing an adaptive learning and intelligent environment based on an identification of students' profiles.

Keywords: component; learning styles, metacognition, data analysis, PCA, learning environment

1 Introduction

The term metacognition was first brought to the literature by J.H Flavell in his research on meta memory functioning [1], it is defined in a large quantities of research papers, as thinking about thinking, or cognition about cognition. It takes place on every learning process, it allows students to control and monitor their own learning, such as selecting the appropriate strategies, intervening, monitoring the execution and even evaluating its effectiveness. Metacognition is a conscious processes, according to many authors [2], [3],[4] metacognition is described as the second order level of cognition.

According to [5], metacognition is a state of awareness and knowledge of one's mental processes such that one can monitor, regulate and direct them to a desired goal.

Livingstone [6] suggests that metacognition refers to higher order thinking, that involves active control over the cognition process which influences the learning success significantly, because students with high metacognitive and self-regulatory abilities are more successful. Metacognitive skills involve the process of individuals deciding what strategy to use in what situations as a result of the metacognitive experiences

DOI: 10.14738/tmlai.54.2968
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2968

they have gone through, using a strategy, monitoring learning, changing that strategy if learning has not taken place and trying a new one.

We summarize all the ideas, and confirm that metacognition is an important element of cognition, because it's about students knowing and progressing, it helps students retain more information and guides them on how to proceed to be more successful to increase achievement [7].

This positive role of metacognition in learning - and comprehension activity is supported by three theories of learning; cognitive development, behavioristic and information processing learning theories [8]. The role is illustrated by a series of papers and experimental studies to assess the impact of metacognitive skills on learning performance across many disciplines; therefore, our objective is, firstly, to examine the relationship between metacognition and learning styles among university students. Secondly, we tend to classify the sample of students according to their learning styles and metacognitive abilities in reading and comprehension activities.

In the next part of the paper, we expose metacognition more deeply, we will present the nature of reading for comprehension, we will see what is the characteristics of good reader, and we will highlights relations between learning styles and control and awareness.

2 Metacognition while reading – comprehension activities

Students use a variety of strategies to work with texts, they take actions to deal with information in the text, and to manage and monitor their understanding intentionally and carefully. While reading, students sometimes need to apply some problem solving strategies to clear up misunderstandings that could arise; on the other hand, some strategies are used as basic mechanisms to aid reading comprehension, for example: note taking, underlining and highlighting textual information, summarizing, etc...Therefore, reading is more than decoding print, it is not just a basic skill, and it is a complex process that refers to a particular world of knowledge and experience.

We are going to underline the relationships between reading for understanding and metacognition defined earlier; we present our conception of the nature of reading as a brief outline of what we have learned from recent research and observations.

2.1 Nature of reading for comprehension

Reading is a process; the major product consists of creating a mental representation of the text, which serves as an evolving reference for understanding subsequent parts of the text. As the students read further, they test every meaning and start to monitor their understanding; at every moment they can draw a variety of strategies to readjust their progress, it is about taking distance between oneself and the text content.

Reading is not only words and sentences on the paper, but also the ideas, memories, and personal knowledge evoked by those words and sentences, furthermore, a person who understands one genre of a text is not necessarily proficient at all genres, a chemistry teacher may feel completely insecure when trying to understand some of the original source history materials, while a political science undergraduate can understand more and better. In other words, reading is a complex process, influenced by multiple and situational factors, and needs to be controlled and managed s

2.2 Characteristics of proficient readers

Different research papers about reading describes some characteristics of good or proficient readers, they possess a set of key habits which could be summarized as follows according to [9].

- Motivated to read and to learn,
- Mentally engaged,
- Socially active around reading tasks,
- Strategic in monitoring the interactive processes that assist comprehension:
  - ✓ Setting a goal that shapes their reading processes,
  - ✓ Monitoring their emerging understanding of a text, and
  - ✓ Coordinating a variety of comprehension strategies to control the reading process.

To that end, the awareness level of the students can be predicted by the use of metacognitive reading strategies, which can inform us about the reading proficiencies, according to [10].

Novice reader's score lower than good readers in using all reading strategies, especially some sophisticated cognitive and metacognitive strategies, because they (novice readers) seem often oblivious to these strategies and the need to use them.

2.3 Metacognitive strategies in reading for comprehension

Metacognitive strategies are defined by [11] as: Behaviors undertaken by the learners to plan, arrange, and evaluate their own learning, such strategies include directed attention and self-evaluation, organization, setting goals and objectives, seeking practice opportunities, and so forth.

In the context of reading, self-monitoring and correction of errors are further examples of metacognitive strategies.

A classification of reading strategies is presented by Oxford [12], in line with [11] and [13], they are defined as:

- Cognitive reading strategies used to manipulate the language that includes note taking, summarizing, paraphrasing, predicting, analyzing and using context clues.
- Memory reading strategies, which are techniques used to assist the learner to recall information, such as word association and semantic mapping.
- Compensation reading strategies, such as 'inferencing', and guessing while reading, which can assist the learner in making up for reading deficiencies.
- Affective reading strategies, which include self encouraging behavior to lower anxiety, such as rewarding oneself for reading efficiently.
- Social reading strategies, involving collaborating with peers, for example, to ask questions, seek help or correction and to get feedback while reading.

It's evident that the use of these strategies deals with an additional factor of learning; because every student is different, every student has his own way in which he approaches and responds to a learning experience according to his own learning style. In general, metacognition is the engine that drives self directed learning [14].

3 Learning styles and metacognition

Learning styles refers to different approaches or ways of learning, students take in information in different ways, by seeing and hearing, reflecting, and acting, reasoning logically and intuitively analyzing and visualizing, therefore our teaching methods also vary from one another. When mismatches exist between learning styles and the teaching manner used in class, or the pedagogy employed in the learning environment, the students may become bored, inattentive, get discouraged, and in some cases abandon the task of learning altogether.

To overcome these problems, a differentiation should take place in a way that all students learn in a manner they prefer according to their style of learning, this can increase comfort and willingness to progress and learn. Recent research claims that learning styles can hurt learning; we will discover those results in discussion section to show you how it is suggested to deal with learning styles.

3.1 Assessment tools of learning styles

Using a variety of assessment tools, individuals can discover their own interest levels for a set of criteria to help establish the methods in which they could obtain much information and improve their learning, one assessment tool that can be used in establishing a person's learning style is the index of learning styles (ILS), the ILS result provides an indication of learning preferences, possible strengths and tendencies or habits that might lead to difficulty. The instrument used to assess preferences on four dimensions (active/reflective, sensing/intuitive, visual/verbal, and sequential/global) of the learning style model developed by Richard M. Felder and Barbara A. Soloman of North Carolina State University [15].

Table 1. Description of learning style scales

• Active : Learn best by doing something active with the information they are learning
• Reflective : Reflective and prefer to think about the information first
• Sensitive : Sensing like learning facts
• Intuitive : Prefer discovering possibilities and relationship
• Visual : Remember best what they see
• Verbal: Get more out of written and spoken information
• Sequential : Gain understanding in linear, logical steps
• Global : Tend to absorb material almost randomly without seeing connections

3.2 Relationship between metacognition and learning styles

One of the main difficulties that student faces while trying to develop their understanding is an overall lack of awareness to their own learning process [1]. Flavell describes three basic types of awareness related to metacognitive knowledge, the first one is an awareness of knowledge, which is described as an understanding of what the student does and does not know, and what he wants to know, the second one is an awareness of thinking, which describes her understanding of cognitive tasks and the nature of what is required to complete them successfully, finally, there is an awareness of thinking strategies, defined as an understanding of approaches and strategies to regulate learning progress.

Focusing on the function of metacognition itself, there is agreement upon its operational definition. It is a manifestation of the monitoring and regulatory function.

Efklides[16] presented three facets of metacognition distinguished by their manifestations as a function of monitoring and control.

Monitoring functions are metacognitive knowledge, knowledge about one's cognition, metacognitive experiences, and metacognitive judgment and assessment which have a very significant impact on learning experience, the control function of metacognition is metacognitive skills; it's about the knowledge of the procedures needed to control cognition. However, the existing literature has paid less attention to the sources of metacognition; the starting points of the construction of metacognition. As mentioned above, individual differences play a critical role because students learn differently and possess differing conceptual systems according to their learning styles and approaches.

4 Significance of the study

The basic goal of this study is to determine the relationships between metacognition and learning styles of the students attending the University of Science and Technology. Additionally, the final objective of our study on metacognition is to find how an ideal learning environment could support metacognition in learning as an important dimension of effective learning.

4.1 Materials and method

- Population and sample of the study

A total of 80 University students were involved in this study, data were collected in the first hand to measure the student's metacognition abilities in reading and comprehension as well as to determine their learning styles according to ILS model.

- Instruments

Two survey instruments were used in this study, each is described as follows:

4.1.1 Survey of reading strategies

The Survey of Reading Strategies Questionnaire (SORS) was developed by Mokhtari and Sheorey (2002)[17] and used to collect data, the questionnaire was used so that the students could indicate the extent to which they used metacognitive reading strategies.

The SORS comprised three subscales: Global Reading Strategies or GLOB, Problem Solving Strategies or PROB, and Support Reading Strategies or SUP. According to Mokhtari and Sheorey (2002)[17], students use Global Reading Strategies to work with text directly or to manage and monitor their reading intentionally and carefully.

Problem Solving Strategies are used for solving problems of understanding that arise during the reading of a text. Support Reading Strategies are used as basic mechanisms intended to aid reading comprehension, for example through note-taking, underlining and highlighting textual information (Mokhtari & Sheorey 2002)[17].

SORS has demonstrated reliability and validity, the internal consistency reliability coefficient (as determined by Cronbach's alpha) was 0.92 for the Global Reading Strategies, 0.79 for the Problem Solving Strategies, and 0.87 for the Problem Solving Strategies.

The researcher administered the questionnaires to his classes as indicated earlier, the students were asked to circle the number that applied to them indicating the frequency with which they used the reading strategy described in the statement, the students took a maximum of 20 minutes to complete the questionnaire in class under the researcher's supervision, clarification of questions were made by the researcher where and when necessary.

4.1.2 Index of learning styles questionnaire (lsq)

The Felder and Silverman's Index of learning styles is one of the most widely used models of learning styles, in this model, we distinguish four dimensions of learning styles, each dimension is a continuum with one learning preference on the far right and the other on the far left.

The results of the questionnaire provide an indication of individuals' learning preferences and an even better indication of the preference profile of a group of students.

The instrument consisted of 44 items that relate to learning styles, for each of the 44 questions, the student has to choose only one answer from two for each question, which seems to apply to her more frequently.

Table 1.Description oft he contentof LSQ

Active /Reflective	11 items	How the student prefer to process information
Sensing/ Intuitive	11 items	How the student prefer to take in information
Visual/ Verbal	11 items	How the student prefer information to be presented
Sequential/ Global	11 items	How student prefer to organize information

Data Collection

The participants were given oral description of the objectives of the study and that their response would be used only for research purposes; participants were given 30-35 minutes to respond, and they were motivated to take the surveys seriously, and were ensured that the obtained data would be kept confidential.

4.1.3 Data Analysis and results

Data of metacognition survey and learning style were computed separately with the help of Excel, and then Minitab.17 was employed to compute the data gained in the study, we use the statistical procedure of data analysis for testing the relationship between metacognition and learning styles.

Our interest is to see if there is a relationship between student's metacognitive level and their learning styles.

Distribution according to Learning Styles

On the ACT/REF scale we observe that 40,5% of students are well balanced, which means they learn affectively either by doing something active with information, for example discussing, or applying it or explaining it to others, they prefer to think about information quietly, in another hand 31,6% have tendencies to action, they feel retain more information when they find ways to do something with it, 19% of students are reflective so they could retain the materials more effectively if they find ways to reflect

upon them and review what they learn and think about it, finally only 3,8% are Strongly Active, while 5,1% are strongly reflective.

On the sensitive/Intuitive scale, students are sensing learners, 45,6% are sensitive, 10,1% are strongly sensitive, they prefer to see connections between information and real world, they don't get the whole of the material if it deals only with abstract details or theoretical data. In another hand 32,9% are well balanced, while 10,1% are intuitors , only 1,3% of students are aimed to be strongly intuitors, they have the preference to learn by discovering possibilities and relationships, in general they may have trouble with boredom in situation that deals with memorization.

We observe that 36,7% of our population have preference to visual learning style, the same percentage are strongly visual learners, they learn more effectively with any type of visual representation of course material, while only 5,1%, and 1,3% aimed to be respectively verbal and strongly verbal, they get more out of words, written and spoken explanations.

On the scale of sequential and global learning style, we have observed that 36,7% are well balanced in that scale, 13,9% are aimed to prefer Global learning style, while 1,3% of the population is very global, they tend to learn in large jumps, learn almost randomly without seeing connections. 38% of students prefer the sequential style, they tend to learn in linear steps, the more they put the material in a logical path the more they feel comfortable and the best they learn.

We confirm that everybody is for example, active sometimes and reflective sometimes, the same for the other scale, but the strong or the moderated preference for one or the other which highlights some cautions, if the student overemphasize intuition he may miss important details or make some mistakes, if he overemphasize sensing he may rely too much on memorization and not concentrate enough on critical and innovative thinking.

Principal component analysis

In order to get knowledge of our experimental data, we choose to perform dimensionality reduction of the space, while preserving as much of the randomness in the high dimensional space as possible [18], it is a basic idea behind Principal component analysis, here we take the cloud of observations or points, and rotates to make visible the maximum variability.

PCA is a data analysis technique to identify relations between features; we use it to extract and select features and to find response related to our research guiding questions.

The task of investigation is a feature extraction and selection method, the goal is to obtain the most relevant information from the observation and features regardinglearning styles and metacognitive awareness and represent that information in a lower dimensionality space, our questions concerns the relations between the way students prefer to learn, as learning styles, and the awareness in learning as metacognition, to discover important features of data set, and interpret to construct a useful analysis.

We conduct our analysis on a multidimensional space of data representing the profiles of learners and their metacognitive skills, we will highlight in the next part of the paper the relationships between styles in learning and Metacognitive problem solving skills, then in the discussion we will provide you with further details.

4.1.4 Learning styles and Metacognitive Problem Solving Strategies.

The eigenvalues indicate that three components provide a reasonable summary of the original information, accounting for 70% of the total variance.

The table of principal components (Table 5) reveals a strong correlation between Problem solving strategies variable and being sequential, active and sensitive,(SN 0.414 SEQ 0.266) on the principal component one, but this kind of skills is affected positively with reflection (RF -0.583 M_MetaPRO -0.493).

Table 2.Tableofprincipalcomponentsanalysis

Variable	PC1	PC2	PC3
AC	0,202	0,583	-0,058
RF	-0,181	-0,608	0,072
SN	0,414	0,143	0,058
IN	-0,414	-0,139	-0,053
VI	-0,018	0,046	-0,624
VR	0,014	-0,047	0,623
SEQ	0,266	-0,039	0,267
GL	-0,303	0,041	-0,238
M_MetaPRO	0,649	-0,493	-0,284

The first component will follow that students with large sense of sensitivity in learning, and also logical information processing, whereas communities with small values would have very few of these types of characteristics (Fig.6).

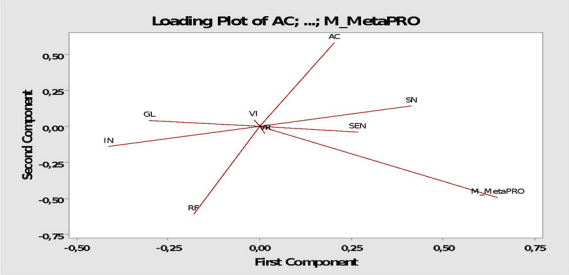


Figure 1. Loading plot of learning styles and Problem solving strategy

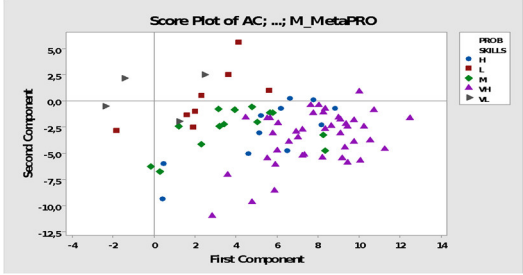


Figure 2. Score plot in dicating level of mastery of Problem solving strategy

To complete the analysis of learning styles and problem solving skills, we produce a scatter plot of the component scores, where we plot the second against the first component, (see Figure.1) each dot represent one student, we give them five kind of symbols by degree of proficiency in problem solving skills, in this present context, we wish to identify the location of the points to see if students with high levels of given component tend to be clustered in a particular manner or a particular region of the plot. Effectively student with high proficiency are situated in the region with high value for the first component (VH, H, M), the student with grade VL-Very Low have a low values, these poor student are more active, less sequential and less sensitive than others.

5 Discussion

In an effort to investigate the relationships between the way learners control and monitor learning process (metacognition), and their preferences as learning styles, we performed this case study, we apply data analysis to explore relations and do useful interpretations using principal component analysis, we overcome the multidimensionality of the dataset, we identify the relations and patterns behind the data of the experimentation.

As conclusion, metacognition is strongly related to learning styles, this is probably due to the importance of the personal identity of learners, each one learn according to some criteria, which may hurts learning in some cases when environment don't enhance those preferences.

5.1 Learning styles affects global thinking strategies

Global Reading strategies (GLOB), which can be thought of as generalized strategies, it aim is setting the stage for reading act, we consider this element as the basic strategies of reading for comprehension, localized at the start point of the learning act, focused on evaluating what to read, what to ignore, setting goal, noting characteristics of the task, guessing, etc.. All this strategies are more or less affected by styles of learning.

The first finding of this research reinforce the idea that Intuition and Global learning preferences affects positively the way every learner control and monitor his learning process; the Visual and the balance between Active and Reflective preferences might help to enhance metacognition, while Sequential style may hurts global strategies. The first component will follow that students who are global, with large sense of intuition in learning, whereas communities with small values would have very few of these types of characteristics, the second component will follow active and visual preferences. As shown in Fig.14,both the second component (Y1) as described in the first finding and the first (X1) affects the level of global strategy; the darker green regions indicate higher quality.

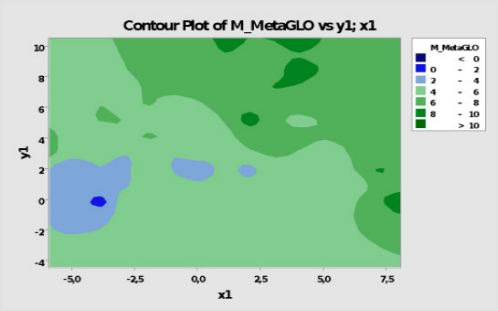


Figure.3: Contour plot describing level obtained on global strategy on a numeric scale

From the definition of intuition, which can be developed with accumulating experience, after working on several situations, our results confirm the importance of this element on making reasonable connections without formalizing the arguments. On other hand global preference came to adjust intuition, her place is not to destroy all intuition, so according to our results on global metacognitive skills, we claim that global thinking should be used to destroy bad outcomes of intuition, while clarifying and elevating good intuitions. We propose that only good combinations of both global processing and intuition that can lead

learner to tackle learning problems, the first (Global style) helps to deal with the big picture while the latter (Intuitive style) helps to deal with the fine details.

5.1.1 Learning styles are the keys to enhance problem solving strategies

Problem solving strategies are used when problems arises, known as repair strategies they enable learners to overcome trouble, control understanding, for example in case of losing concentration, learner may decide to go back for better understanding.

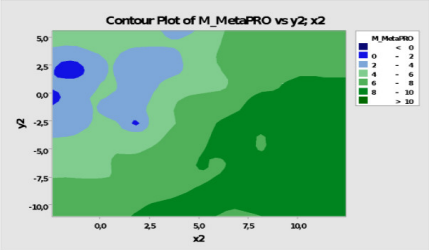


Figure. 1:Contour plot describing level obtained on problem solving strategy on a numeric scale

For this element we confirm that z Reflective, Sequential and Sensitive learning styles are the best styles to boost problem-solving strategies.

The first component (X2) will follow that students with large sense of sensitivity in learning, and also logical information processing, whereas communities with small values would have very few of these types of characteristics, the following contour plot (Fig.15) explores the potential relationship between the two components and the problem solving skills.

We have seen that problem solving strategy is related to both sensitivity and reflection as learning styles, this results ally with findings related to this type of strategy in learning, since sensing is considered as the ability to learn by manipulating objects, skills about learning inductively rather than deductively. In general sensitive learners tend toward psychomotor over abstract thinking, they prefer personal connections to topics, they follow directions they have written themselves, and they benefit more from demonstrations. The second element is reflection, enables learners to reflect upon their results and outcomes, in order to adjust and monitor eventual future actions, we suggest that getting best performance in problem solving strategy is enhanced by fostering sensing and reflection in learning.

5.2 Support strategies are influenced by learning styles

Support strategies are the third broad subcategory measured, it consists of providing learners with additional support mechanism, the third finding of this research confirms the idea that SENSITIVE, ACTIVE, VERBAL learning preferences following the first component (X3), has all positive influence on the way learners search for additional support, using Support strategies.

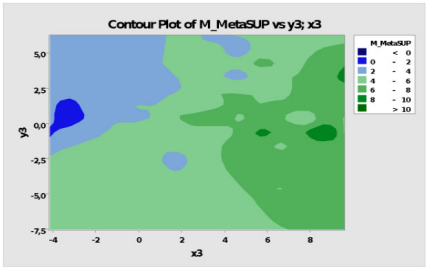


Figure. 2: Contour plot describing level obtained on support strategy on a numeric scale

Support strategy can be enhanced by fostering sensitivity in learning, active processing and verbal instructions, effectively learner benefits from those preferences to enhance their level of proficiency, support reading strategies are used as basic mechanisms intended to aid reading comprehension such as note-taking, underlining and highlighting textual information, those qualities might be enhanced with some aids such as collaboration, concept map, and feedbacks.

6 Conclusion

In the previous part we present a summary of our findings, and explanation of how it confirms what was described in the literature, the present study suggests that understanding learning styles of learners might contribute to increase the efficiency and quality of learning, good identification of each one's preferences, weaknesses and strengths is the key to make effective learning happen, we have used principal component analysis as a tool to apply an exploratory data analysis and predictive analysis, to explain the variation in data, it allows us to see how different variable change with each other, and enable us to select a minimum number of principal components to explain most of the variation and to select which variables are the most significant in describing the full data set. This study may provide educational with good insights about cognitive process involved in learning, and what differentiates successful from less successful students. Because understanding begins to develop when students have frequent opportunities to deploy some strategies to control and monitor the reading process, this ability can be enhanced as they have more opportunities, more support, and encouragement.

Actually we work on several implications for instructional interventions in e-learning context, our focus is to teach students how to be more aware of their learning processes and products, as well as how to monitor and regulate those processes for an effective learning. We believe that building intelligent environment for human learning, with additional care for styles of learning, motivation and individualization can permit advances in the area of metacognition and e-learning.

REFERENCES

[1] J. H. Flavell, 1979. Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. American Psychologist, 34(10), 906-911.

[2] Nelson & Narens, 1994. Why investigate metacognition ? In J.Metcalfe & AP. Shimamura (Eds). Metacognition: knowing about knowing (pp. 1-26). Cambridge, MA: MIT Press

[3] Schraw & Dennison, 1994. Assessing metacognition awareness, Contemporary Educational Psychology,19, 460-475

[4] Pintrich, & Zeidner, 2000. The role of goal orientation in self-regulated learning. Handbook of self-regulation (pp, 452-494), San Diego, CA: Academic Press

[5] Harris, T.L, & Hodges, R.E, 1995 . The literacy dictionary: The vocabulary of reading and writing. Newark, DE: internationai Reading Association

[6] Livingstone .J.A, 1996. Effects of metacognitive instruction on strategy use of college students. Unpublished manuscript, State University of New York at Buffalo.

[7] Caliskan, M., & Sunbul, A, 2011. The effects of learning strategies instruction on meta-cognitive knowledge, using meta-cognitive skills and academic achievement, primary education sixth grade Turkish course sample. Educational Sciences: Theory & Practice, 11(1), 148-153

[8] Hoskin, P. W. O, 2000. Affecting increased student achievement in geoscience education by instruction in metacognition: A small class case study. Retrieved from http://www.usq.edu. au/electpub/ejist/docs/Vol7_no2/ CurrentPractice/Affect_ student_ach.htm

[9] Baumann & Duffy, 1997 Engaged Reading for Pleasure and Learning: A Report from the National Reading Research Center.

[10] Downing, K., Kwong, T., Chan, S., Lam, T. and Downing, W, 2009. Problem-based learning and the development of metacognition. Higher Education, 57 (5), 609-621.

[11] Singhal, 2001. Reading proficiency, reading strategies, metacognitive awareness and L2 readers. Reading Matrix, 1-9

[12] Oxford, R. L, 1990. Language learning strategies: What every teacher should know. USA: Heinle & Heinle

[13] Bianco, M, 2010. La compréhension des textes : peut-on l'apprendre et l'enseigner ? In M. Crahay & M. Dutrevis (Eds.), Psychologie des apprentissages scolaires, (pp. 229-256). Bruxelles: De Boeck.

[14] Nelson, S., & Conner, C, 2008. Developing self-directed learners. Retrieved January 15, 2008

[15] R. Felder and J.Spurlin. "Applications, Reliability, and Validity of the Index of Learning Styles,"

[16] Efklides, 2006. Metacognition and affect: What can metacognitive experiences tell us about the learning process?. Educational Research Review, 1, 3-14

[17] Mokhtari and Sheorey, 2001. Differences in the metacognitive awareness of reading strategies among native and non-native readers. System, 29, 431-449.

[18] I.T.Jolliffe. Principal component analysis, Second Edition, Springer, 2nd ed. P.cm – Springer series in statistics, ISBN 0-387-95442-2

[19] Jackson, J.E . 1991, A user's guide to Principal Components. New York : Wiley.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Improved HMM for Cursive Arabic Handwriting Recognition System Using MLP Classifier

[1]Mouhcine Rabi, [2]Mustapha Amrouch, [3]Zouhir Mahani
[1,2]Laboratory IRF-SIC, faculty of sciences IbnZohr University, Agadir Morocco
[3]High school of technology, IbnZohr University, Agadir Morocco
mouhcineh@gmail.com, m.amrouch@uiz.ac.ma, zouhir.mahani@uiz.ac.ma

ABSTRACT

Recognizing unconstrained cursive Arabic handwritten text is a very challenging task the use of hybrid classification to take advantage of the strong modeling of Hidden Markov Models (HMM) and the large capacity of discrimination related to Multilayer Perceptron (MLP) is a very important component in recognition systems.The proposed work reports an effective method on improvement our previous work that takes into consideration the context of character by applying an embedded training based HMMs this HMM is enhanced by an Artificial neural network that are incorporated into the process of classification to estimate the emission probabilities. The experiments are done on the same benchmark IFN/ENIT database of our previous work to compare the results and show the effectiveness of hybrid classifier for enhancing the recognition rate the results are promising and encouraging.

Keywords: Arabic Handwriting Recognition; Context; Embedded training; HMMs; Multilayer Perceptron (MLP).

1 Introduction

Systems for handwriting recognition are referred to as off-line or on-line systems depending on whether ordinary handwriting on paper is scanned and digitized or a special stylus and a pressure-sensitive tablet are used. In both the ultimate objective is to convert handwritten sentences or phrases words or characters in analogue form (off-line or on-line sources) into digital form (ASCII).

More than 300 million people around the world speak Arabic and their derivative. Arabic is naturally written cursively in both handwritten and typewritten modes. In comparison to Latin Arabic seem to be more complex. For example many letters in this language have complementary diacritics such as dots madda and zigzag bars. In addition the letters have different shapes at different locations of the word.

Due to variability in handwriting styles and distortions caused by the digitizing process even the best handwritten word recognizer is unreliable when the number of word choices is large. This necessitates the use of advanced concepts to achieve a performance level comparable to that of humans. The researches focus on the use of new methods and approaches to perform handwriting recognition. In this area the concept of combining multiple classifiers is proposed as a new direction for the development of highly reliable handwriting recognition systems and some preliminary results have indicated that the

DOI: 10.14738/tmlai.54.2969
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2969

combination of several complementary classifiers will improve the performance of individual classifiers [01] [02].

In current work our system performs training and recognition of words and characters. In order to model the variations related to the character context in the corpus we have opted for a specific type of learning. Therefore character models are trained and obtained through embedded training; thereafter the decision has been done by the proposed hybrid classifier which mainly based on HMMs and neural network.

The remainder of this paper is organized as follow. Section 2 presents a literature review and related works of handwriting recognition system. Section 3 is focused on our contribution starting with our developed reference system then the incorporation of hybrid classifier. The performance of the recognition system has been experimented on the benchmark database IFN/ENIT and the obtained experimental results are shown and analysed in section 4. The paper finally closed with a conclusion and perspectives.

2 Literature review

Hidden Markov Models (HMMs) have proven to be one of the most successful and widely used classifiers in the area of text recognition. There are many reasons for success of HMMs in text recognition including avoidance of the need to explicitly segment the text into recognition units; characters or graphemes In addition HMMs have sound mathematical and theoretical foundation [03].

[04] The authors have investigated on contextual sub-characters HMMs for text recognition by using multi-stream HMMs where the features calculated from a sliding window frame form one stream and its derivative features are part of the second stream. The experiments were conducted with different train-test configurations on the IFN/ENIT database and the best recognition rate achieved was 85.12%. In [05]Azeemused an effective technique for the recognition of offline Arabic handwritten words using Hidden Markov Models. Besides the vertical sliding window two slanted sliding windows are used to extract the features. Three different HMMs are used: one for the vertical sliding window and two for slanted windows a fusion scheme is used to combine the three HMMs. [06] proposed a combined scheme for Arabic handwritten word recognition using a HMM classifier followed by re-ranking. Basically intensity features are used to train the HMM and topological features are used for re-ranking to improve accuracy. Experiments were carried out using IFN/ENIT database and the achieved recognition rate was 83.55%.An alternative approach is proposed in [07] and [08] where multi-stream HMM models is used this paradigm provides an interesting framework for the integration of multiple source of information. Significant experiments have been carried out on two public available database the recognition rate was 79.8% in IFN/ENIT for Arabic and 89.8% in IRONFF for Latin script. In [08] Maqqorproposed a system of Arabic handwriting recognition based on combining methods of decision fusion approach. The combination of the multiple HMMs classifiers was applied by using the different methods of decision fusion approach. The system is evaluated using the IFN/ENIT database. Experimental results demonstrate that the Weighted Majority Voting (WMV) combination method have given better recognition rate 76.54% with Gaussian distribution. In [09] the authors present an analytical approach of an offline handwritten Arabic text recognition system. It is based on the Hidden Markov Models (HMM) Toolkit (HTK) without explicit segmentation.The feature extraction uses a sliding window on the line text image and processed by two groups of these features (the features of local densities and the statistical characteristics).The proposed system has been experimented in two different databases: “Arabic-Numbers” database where we

achieved a rate of 80.26% for words and 37.93% for sentences and IFN/ENIT database where we achieved a rate of 78.95% for words.

Otherwise [10] proposed a new approach for the offline Arabic handwritten word recognition based on the Dynamic Hierarchical Bayesian Network (DHBN) using a free segmentation released by a smoothed vertical projection histogram with different width values. The model is consisting of three levels. The first level represents the layer of the hidden node which models the character class. The second layer models a frame set representing the sub-characters and the third layer models the observation nodes. The developed system has been experimented and the results are provided on a subset of the IFN/ENIT benchmark data base. These results show a significant improvement in the recognition rate because of the use of the DHBN. Most of the recognition errors of the proposed system can be attributed to the segmentation process error and to the poor quality of some data samples.

In many other works neural networks in its different applications have been extensively applied to classify characters as part of isolated or continuous handwritten word recognizers [11] [12] [13] [14] [15].

This paper focus on the impact of using aembedded training based on a hybrid classifier the motivation for the work on the hybrid HMMs and Artificial neural network models presented here originates from a critical analysis of the state of the art in offline handwritten text recognition [16] [17] [18] our previous work on offline handwriting recognition using HMMs [19] researches and experiences in using hybrid HMM/ANN models for automatic speech recognition [20][21][22][23] [24]and for online handwriting recognition [25]. All these criteria making hybrid modeling an important factor in order to achieve an effective and efficient system.

3 Contribution

3.1 Reference System [19]

Our reference system (figure1)was analytical without explicit segmentation based HMMs using embedded training to perform and enhance the character models.

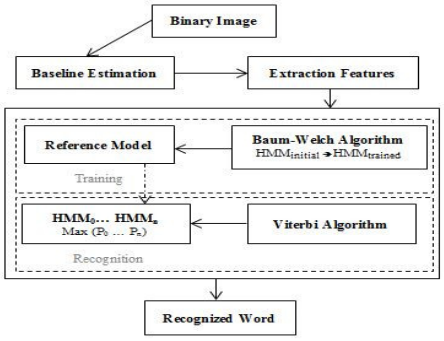


Figure.1. Synopsis of reference system

Extraction features was preceded by baseline estimation; the approach used to estimate these baseline based on the horizontal projection curve that is computed with respect to the horizontal pixel density knowing that the skew and slant correction of words are made in pre-processing step to harmonize the direction of the sliding windows in the extraction features. These latter are statistical and geometric to integrate both the peculiarities of the text and the pixel distribution characteristics in the word image.

The sliding windows are shifted in the direction of writing (right to left). In each window we extract a set of 28 features represent the distribution features based on foreground pixels densities and concavity features. Each window is divided into a fixed number n of cells. Some of these features are extracted from specific areas of the image delimited by the word baselines.

These features are modelled using hidden Markov models and trained by theembedded training method (figure 2).

We used a model for each character right-left topology with four states and three transitions for each state. Word model is built by concatenating the appropriate character models.

The embedded training is to automatically identify relevant information letters without specifying them explicitly by exploiting the redundancy of information between words matched to changes in context and letters position.

The major problem of HMMs is the estimation of emission probabilities; this confirms that HMMs are powerful to model sequences but still limited compared to NN and SVM in classification [26] for this reason and the motivations cited above in section II it is prominent and promising to use a hybrid classifier.

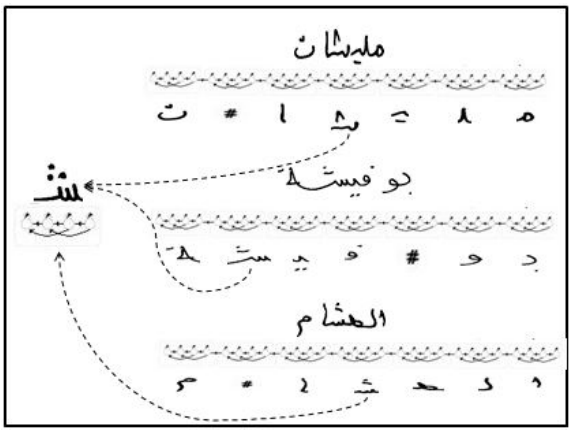


Figure.2. Embedded training of character "chin" [19]

For more explanations about the baseline system refer to [19].

3.2 Hybrid classifier

To improve the performance of the off-line handwritten recognition system either the accuracy ofthe classifier has to be increased. In this section we introduce the hybrid approach then we clarify the principle for our offline Arabic handwriting recognition system.

While HMMs are effective in modeling variation in handwriting they lack discrimination ability because of maximum-likelihood parameter estimation criteria. The strength of MLP is in the fact that they don't need to assume about statistical distribution of input as well as they can be trained to exhibit discriminant properties. As already mentioned recent works in various area of research tried to develop hybrid HMM/MLP systems in which MLPs are used to compute the emission probability associated with each state of HMM.

By HMM the goal of handwriting recognition is to retrieve the most likely grapheme character or word sequence W given a sequence of observation vectors x which is achieved by maximizing the a posteriori probability:

$$\hat{w} = \arg\max P(w \mid X) \tag{1}$$

Typically each grapheme is modelled by a right to left HMM and the number of states is chosenglobally or individually for each character. Gaussianmixtures are used to model the output distributions ineach state q given the feature vector xP(x|q). The Baum-Welch algorithm is used for training the HMMs whereasthe Viterbi algorithm is used for recognition.

Hybrid modelsfor handwriting recognition based HMMs were built with different neural networks. [27]; [28] use an MLP. [29] built an hybrid CNN/HMM. In [30]CNNs is applied in the hybrid framework for handwritten word recognition using different segmentationmethods.

In Hybrid HMM/MLP classifier neural networks can be considered statistical classifiers under certain conditions by supplying output of a posteriori probabilities. Thus it is interesting to combine the respective capacities of the HMM and the MLP for a new efficient recognition system inspired by the two formalisms.

The principal idea behind the MLPNN/HMM hybrid approach as illustrated in Figure 3 is to estimate the output probability density function of each state of the used HMM by the output nodes of the MLP classifier which received features as input. These input vectors are pre-processed to finally estimate the posteriori probability deciding whether the input vector belongs to the desired character class. The MLPNN's output weighted by the priori probability of each class forms the probability density function used for every state of the HMM.

In the hidden Markov modeling approach the emission probability density P(x|q) must be estimated for each state q of the Markov chains that is the probability of the observed feature vector x given the hypothesized state q of the model.

In the proposed hybrid HMM/ANN approach the emission probabilities are provided with a neural network since ANNs can be trained to estimate probabilities that are related to these emission probabilities. In particular an MLP can be trained to approximate the a posteriori probabilities of states P(q|x) if each MLP output unit is associated with a specific state of the model and if it is trained as a classifier.

The a posteriori probability estimates from the MLPoutputs P(q|x) can be converted to emission probabilitiesP(x|q) by applying Bayes rule:

$$\mathrm{P}(x \mid \mathrm{q}) = \frac{P(q \mid x)P(x)}{P(q)} \tag{2}$$

The class priors P(q) can be estimated from the relativefrequencies of each state from the information produced bya forced Viterbi alignment of the training data. Thus thescaled likelihoods P(x|q)/P(q) can be used as emissionprobabilities in the proposed system since during recognitionthe scaling factor P(x) is a constant for all classes.This allows MLPs to be integrated into hybrid structural connectionist models via a statistical framework.

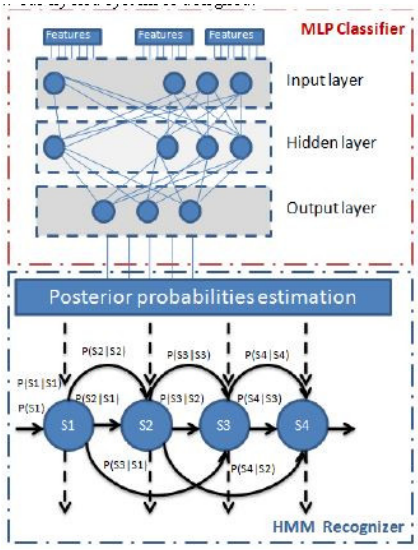


Figure 3: Global scheme design of the hybrid model HMM/MLP

The advantages of this approach are the discriminate training criterion (all MLP parameters are updated in response to every input feature vector) and the fact that it is no longer necessary to assume an a priori distribution of the data. Furthermore if left and right contexts are used at the input of the MLP important contextual information can be incorporated into the probability estimation process.Another strength of this approach is that computing emission probabilities with hybrid HMM/ANN models is usually faster than conventional HMMs with Gaussian emissions since it only requires a forward pass of the MLP for all states of the Markov chains.

4 Experimental results

To evaluate the performance of our recognition system experiments are conducted using IFN/ENIT [31] database of handwritten Arabic words. It was produced by the Institute for Communications Technology at the Technical University of Braunschweig (IFN) and the "EcoleNationaled'ingénieur de Tunis (ENIT) "National school of engineering Tunis."This database is used bymore than 110 research groups in about 35 countries [32].

4.1 HMM vs HMM/MLP Classification

Reference recognition HMM experiments were conducted using continuous density HMMs with diagonal covariance matrices of Gaussians in each state, a right-to-left topology was applied with four states for each character and three transitions for each state. The optical models were trained and tested using the HTK toolkit [33].

The developed system used hybrid classifier based HMMs enhanced by an artificial neural network that are incorporated into the process of classification to estimate the emission probabilities.

Table1 shows the experimental results of our developed system compared to reference system using the same benchmarking database IFN/ENIT to illustrate the reliability of our improvement models.

Table1: Recognition results of improvement system compared to reference system

System	Models	RR* **%**
Reference	HMM	87.93
Hybrid	HMM/MLP	**89.03**

RR: Recognition Rate

We compare the reference system [19] with the developed system using hybrid HMM/MLP classifier. Results in Table1 show an improvement due to embedded training using the hydride classification: accuracy is increased by 1.1%.

4.2 Comparison with other systems

A comparison of the recognition rates of our system with other state-of-the-art systems evaluated on the IFN/ENIT database is presented.

Table2 shows the results of recognition rates for various offline systems recognition of cursive Arabic handwritten text using diverstype of models and the same database with the same configuration (training-testing); sets a, b and c for training and d for test, to compare rates and infer the effectiveness of the proposed method.

Table 2: Recognition results of various systems

System	**Models**	**RR* %**
Irfan [04]	Contextual sub character	85.12
Alkhateeb[06]	HMM+ re-ranking	83.55
Kessentini[07]	Multi-stream HMM	79.80
Maqqor[08]	Multiple classifier	76.54
ElMoubtahij[09]	HMM	78.95
Khaoula[10]	DBN	82.00
Our system	**HMM/MLP**	**89.03**

As it can be noted from Table 2 most of the previous systems are based on HMM using various techniques exploiting the contextual approach, multiple HMM classifier, re-ranking or other techniques to improve HMM models and the recognition rate for the results of the systems mentioned does not exceed 86%. Others used hybrid classifier such as HMM and Dynamic Bayesian Network and the recognition rate achieved was 82.00%. Whereas the proposed system using HMM/MLP outperforms the results and achieve 89.03% due to enhancement of HMMs by incorporating anartificial neural network into the process of classification to estimate the emission probabilities.. This illustrates the effectiveness of embedded training to take account the context of characters to perform the models and using hybrid HMM/MLP classifier to improve the performance of recognition system.

5 Conclusion and perspectives

In this paper we have enhanced the HMM based reference system by using a hybrid HMM/MLP classifier. Extracted features are statistical and geometric to integrate both the peculiarities of the text and the pixel

distribution characteristics in the word image. These features are modelled using hidden Markov models. These models as already mentioned in [19] take into account the context of character by applying aembedded training to perform the models. In addition, the contribution in this paper is the improving of HMM modeling by incorporating MLPs to estimate emission probabilities that present the major HMM problem in order to take advantage of the strength of HMM modeling and neural networks classification.The modelling proposed has improved recognition and shown encouraging results to be perfect using annexes improvements.

Due to variability in handwriting styles even the best handwritten word recognizer is unreliable when the number of word choices is large. This forced the use of linguistic constraints to enhance HMMs modeling by a statistical language model that are incorporated as a post-processing into the process of recognition.

Statistical Language Modeling involves attempts to capture regularities of natural language in order to improve the performance of various natural language applications, e.g. , Speech recognition, Machine translation, Handwriting recognition, Information retrieval and other applications.

The goal of Statistical Language Modeling is to build a statistical language model that can estimate the distribution of natural language as accurate as possible, which could improve significantly the results especially when we extend our system for line and paragraph recognition.

REFERENCES

[1] J. H. Alkhateeb O. Pauplin J. Ren J. Jiang Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition In Knowledge-Based Systems 24 (2011) pp.680-688.

[2] Salvador Espan˜a-Boquera Maria Jose Castro-Bleda Jorge Gorbe-Moya and Francisco Zamora-Martinez “Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN models” 2011

T.Plots and G.A.Fink; “Markov models for offline handwriting recognition: a survey” IntJ.Anal.Recongitvol 12 no 4 pp 269-298 2009.

[3] Irfan Ahmad GernotA.Fink and SabriA.Mahmoud “Improvement in Sub-character HMM Model Based Arabic Text Recognition” 2014 14th International Conference on frontiers in Handwrting Recognition.

[4] S. Azeem and H. Ahmed “Effective technique for the recognition of offline Arabic handwritten words using hidden Markov models” Int. J. Doc. Anal.Recognit.vol. 16 no. 4 pp.399–412 2013.

[5] J. AlKhateeb H. RenJinchang Jiang Jianmin Al-MuhtasebHusni Offline handwritten arabic cursive text recognition using hidden markov models and re-ranking Pattern Recognition Lett. 32 (8) 1081–1088 2011.

[6] Y. Kessentini T. Paquet A. Ben Hamadou Off-line handwritten word recognition using multistream hidden Markov models in: Pattern Recognition Letters 31 (2010) pp. 60-70

[7] A. Maqqor A. Halli K. Satori and H. Tairi Off-line recognition Handwriting combination of mutiple classifiers In 3rd International IEEE Colloquium on Information Science and Technology IEEE CIST’14 October 2014.

[8] El MoubtahijHichamHalliAkram Khalid Satori “Using features of local densities statistics and HMM toolkit

(HTK) for offline Arabic handwriting text recognition" 2016

[9] KhaoulaJayech Mohamed Ali Mahjoub and Najoua Ben Amara "Arabic Handwritten Word Recognition based on Dynamic Bayesian Network" 2016.

[10] Berend-Jan van der Zwaag "Handwritten Digit Recognition: A Neural Network Demo" 2016

[11] Xu Chen Convolution Neural Networks for Chinese Handwriting Recognition 2016.

[12] Charlie Tsai "Recognizing Handwritten Japanese Characters Using Deep Convolutional Neural Networks" 2016

[13] ThéodoreBluche "Deep Neural Networks for Large Vocabulary Handwritten Text Recognition" 2015

[14] Ahmed Mahdi ObaidHazem M. El Bakry M.A. Eldosuky IVA.I.Shehab "Handwritten Text Recognition System Based on Neural Network" 2016

[15] AL-Shatnawi M. Atallah AL-SalaimehSafwan AL-Zawaideh Farah Hanna Omar Khairuddin "Offline arabic text recognition an overview" World Comput. Sci. Inform. Technol. J. 1 (5) 184–192 2011.

[16] M.T Parvez and S.A Mahmoud " Offline Arabic handwritten text recognition : A survey " ACM ComputSurvvol 45 no 2 pp 23-35 May 2013.

[17] A.Lawgali : "A Survey on Arabic Character Recognition" International Journal of Signal Processing Image Processing and Pattern Recognition. Vol. 8 No. 2 (2015) pp. 401-426.

[18] M.RabiM.AmrouchZ.MahaniD.Mammass "Recognition of cursive Arabic handwritten text using embedded training based on HMMs" Published in: Engineering & MIS (ICEMIS) International Conference on Sept. 2016 INSPEC Accession Number: 16467172 DOI: 10.1109/ICEMIS.2016.7745330 Publisher: IEEE

[19] Mohamed ETT AOUIL Mohamed LAZAAR Zakariae EN-NAIMANI "A hybrid ANN/HMM models for arabic speech recognition using optimal codebook" 8th International Conference on Intelligent Systems: Theories and Applications (SITA), IEEE 2013.

[20] Supriya S. Surwade "Speech Recognition Using HMM/ANN Hybrid Model" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 6 4154 – 4157.2015

[21] A. I. G-Moral, U. S-Urena, C. P-Moreno and F. D-Maria, 2011, "Data balancing for efficient training of hybrid ANN/HMM automatic speech recognition, " IEEE trans. on audio, speech and lang. proc., Vol 19, No. 3, 468-481.

[22] Anuj Mohamed K.N. Ramachandran Nair "HMM/ANN hybrid model for continuous Malayalam speech recognition" Published by Elsevier Ltd. Selection and/or peer-review under responsibility of ICCTSD 2012 (International Conference on Communication Technology and System Design).

[23] E. Trentin, M. Gori, 2003, "Robust combination of neural networks and Hidden Markov Models for speech recognition," IEEE trans. on Neural net., Vol 14 , No. 6, 1519-1531.

[24] NajibaTagouguiHoucineBoubakerMonjiKherallah Adel M. ALIMI "A Hybrid NN/HMM Modeling Technique for Online Arabic Handwriting Recognition" International Journal of Computational Linguistics Research Volume 4 Nimber 3 Septembre 2013 pp 107-118.

[25] M.RabiM.AmrouchZ.MahaniD.Mammass" Evaluation of Features Extraction and Classification Techniques for Offline Handwritten Tifinagh Recognition " Global Journal of Computer Science and Technology (USA),: C Software & Data Engineering, Volume 16 Issue 5 Version 1.0 Year 2016.

[26] Dreuw P. Doetsch P. Plahl C. Ney H. (2011). Hierarchical hybrid MLP/HMM or rather MLP features for a discriminatively trained Gaussian HMM: A comparison for offline handwriting recognition Image Processing (ICIP) 18th IEEE International Conference on p. 3541-3544 Sept.

[27] S Espana-Boquera MJ Castro-Bleda J Gorbe-Moya F Zamora-Martinez "Improving offline handwritten text recognition with hybrid HMM/ANN models" IEEE transactions on pattern analysis and machine intelligence 33 (4) 767-779

[28] Q.GuoF.WangJ.LeiD.TuG.Li "Convolutional feature learning and Hybrid CNN-HMM for scene number recognition" Published in journal Neucomputing archive Volume 184 Issue C April 2016 Pages 78-90 Elsevier Science Publishers B.V Amsterdam The Netherlands.

[29] Bluche T. Ney H. &Kermorvant C. (2013b).Tandem HMM with convolutional neural network for handwritten word recognition.In 17th International Conference on Acoustics Speech and Signal Processing (ICASSP) (pp. 2390--2394).IEEE.

[30] M. Pechwitz, S.S. Maddouri, V. Maergner, N. Ellouze, H. Amiri IFN/ENIT – Database of Handwritten Arabic WordsCIFED 2002, Hammamet, Tunisia (2002), pp. 129–136

[31] Märgner V. El Abed H.: ICDAR 2011 - Arabic handwriting recognition competition. In: Int'l Conf. Document Analysis and Recognition pp. 1444–1448 (2011)

[32] S. Young al. the HTK Book V3.4 Cambridge University Press Cambridge UK 2006

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Body Temperature Patterns in Captive Barbary Striped Grass Mousse (Lemniscomysbarbarus): Influence of Photoperiod and Sex Steroids

Malika Nait Daoud, Ali Ouarour
Laboratory of Biology and HealthUniversityAbdelmalekEssaâdi Tétouan, Morocco
Malika-nd@live.com

ABSTRACT

The Barbary striped grass mouse (Lemniscomys Barbarus) has been described as a diurnal and photoperiodic rodent, but no reliable data exist on the daily and seasonal rhythmicity of body temperature (Tb). In this study, Tb patterns were determined in different groups of males maintained in captivity under different regimes of photoperiod. Tb was recorded with surgically implanted iButton temperature loggers. The work shows the existence of a strong and bimodal daily Tb rhythm in this species whose external synchronizer is the light-dark cycle; it's also confirmed the diurnal and photoperiodic character of this animal. The highest values of Tb were registered during the light-phase in the contrary of nocturnal species. This kind of experimental work corresponds to the theme of the ACMLIS'17 conference by means of, on one side, the use of the temperature-sensitive intra-abdominal implants for monitoring core body temperature, an approach of machine learning, and on the other side, the final aim of the work which is to better understand the more complicated and sophisticated of intelligent systems, the central nervous system.

Keywords-body temperature; daily torpor; photoperiod; light-dark cycle; diurnality; Lemniscomys barbarus.

1 Introduction

In this research study, the daily body temperature (T_b) variations following the light/dark cycle were measured on a long time. This kind of experimental works joins perfectly within the framework of the theme of the international conference on the emotional computing, learning machine and the intelligent systems (ACMLIS'17), on one hand, by the use of a technique of continuous measurement of core body temperature, and on the other hand, by their scientific objective to decipher the mechanisms of measurement of the notion of time by the central nervous system. Indeed, no better inspiration for the development of the affective computing as the one who would ensue from a better understanding the functioning of the brain of mammals, a structure of all the human feelings integration. Besides, the technique of the temperature-sensitive intra-abdominal implants for monitoring core body temperature (data loggers DS1922L Thermochron iButtons®) is an approach of the machine learning since it allows an analysis of curves of temporal evolution of a biological measurement, in this particular case the continuous values of the body temperature. So, the definition of the intelligent systems, in a general

DOI: 10.14738/tmlai.54.2970
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2970

framework, covers perfectly the domain of sophisticated interfaces as can well it be the one of the temperature-sensitive intra-abdominal data loggers. The core body temperature-sensitive data loggers are capable of automatically recording temperature over a long defined period of time. They are small, battery powered, and equipped with a microprocessor, internal memories for data storage, and sensors. The sampling methods definition is performed by custom-made computer programs and the data are downloaded to the computer *via* an interface.

Furthermore, in mammals, daily rhythms in many biological processes are a fundamental feature that enables them to anticipate and acclimatize to predictable changes in environment factors. Expression of this rhythmicity with a period of approximately 24-h (*i.e.* circadian) is orchestrated by the suprachiasmatic nucleus (SCN) of the anterior hypothalamus, site of the master endogenous circadian clock [15].

In mammals including humans, like to numerous birds, regular daily T_b variations follow the light/dark cycle. A thermal peak is observed during the day in the diurnal animals and the opposite in the nocturnal ones. These cycles of T_b remain in the absence of any external synchronizer; they are also physiological readout controlled by the SCN [42]. The cost of maintaining high body temperatures in endothermic animals is high. It constitutes a substantial component of their resting metabolic rate and is directly related to ambient temperature [25].

Many small mammals live “right on the edge” with the amount of food that they eat each day barely sufficient to keep them alive and warm until morning. As a result, added stresses on the animal, particularly seasonal shortages of food/water and/or extreme environmental temperatures can be lethal. The solution for many small mammals is to temporarily lower their energy needs by regulating a strong suppression of their metabolic rate, causing T_b fall, and enter either short-term daily torpor or long-term continuous hibernation [18; 20; 33; 9; 2]. Torpor involves the regulation of T_b at a new and substantially lower level, with a new minimum T_b being maintained [8]. Daily torpor lasts for only a matter of hours each day, usually interrupted by periods of diurnal foraging and feeding [19; 43; 12].

Daily T_b rhythm seems to be correlates with the period of activity; the locomotor activity is undoubtedly the best-studied daily rhythmic output of the circadian pacemaker. Concerning this pattern, a difference exists between diurnal and nocturnal organisms because the first ones are active mostly during the day while the second ones are active mostly during the night. The neural mechanisms responsible for diurnality or nocturnality are not well identified [46; 7].

Much less chronobiological studies concern T_b rhythms in diurnal species (e.g., *Octodon degus*, [23]; *Arvicanthis niloticus*, [27]; Arvicanthis *ansorgei*, [4; 41]). These studies, however, are important since diurnal models present similar temporal organization of the sleep/wakefulness rhythm to that of humans and as such, these animal models can be used for suitable biomedical applications. Recently, the Barbary striped grass mouse, *Lemniscomys barbarus*, has been proposed as a new diurnal rodent model for choronobiological studies. More precisely, the circadian organization of wheel-running activity rhythm and cytoarchitecture of the SCN have been described in this African diurnal rodent [21]. We know that *Lemniscomys barbarus* is a photoperiodic species because its transfer from long photoperiod (LP) to short photoperiod (SP) results in a small decrease in its gonadal activity [22].

The objective of the present work was to define better the circadian characteristics of this specific diurnal model by studying the T_b daily rhythms in different photoperiodic regimes. In addition, we investigated

possible role of sex steroids in seasonal changes of daily rhythm of T_b. Moreover, because the recording of T_b was continuous throughout the experiment, we also check if daily torpor or hibernation bouts were expressed during prolonged exposure to SP conditions.

2 Materials and Methods

2.1 Animals

Barbary striped grass mouse, *Lemniscomys barbarus* (Linnaeus 1766) is a rodent (suborder: Sciurognathi; family: Muridae). This species has a strictly African distribution, and can be commonly found in North and sub-Saharan Africa in dry and semi-arid zones, where it frequents local biotopes relatively wet and rich in vegetation. The reproduction of this mammal appears to take place mainly in the spring/summer months (Ouarour, unpublished observations).

Since 2004 a breeding colony is developed at the Faculty of Science in Tétouan (north of Morocco). Animals used in the present study freshly trapped in the region of Tétouan around the village of Benkarrich (around 35°30 N, 5°25 W). At this latitude, photoperiod changes from a maximum of 14h35 to a minimum of 9h43 of light. The annual climate presents mainly two distinct seasons: a raining or wet season from October to April and a drier season from May to September. Ambient temperature oscillates from a minimum during the mild winters of 3-8°C around January to a maximum during the relatively hot summer of 30-35°C around August.

In the area of captures, water is available even in summer. Every year, captures were made throughout the summer (June 15th to September 15th) with traps checked and baited daily with fresh food (bread and olive oil). After transport to the laboratory, trapped animals were housed singly in transparent cages (22x16x14 cm) under controlled squared 16h light-8h dark cycles (LD: 16/8, lights from 04:00 to 20:00 h, white light 200-300 lux during light phase, constant <5 lux red light). Ambient temperature was 22.2±2°C. Water and food (barley and pellets B3, Agropress, Casablanca, Morocco) were supplied *ad libitum*. Animals used in this work had a body weight ranging between 33 and 59 g.

All manipulations of animals were made in Morocco in agreement with local legislation. Moreover, all procedures used in animal experimentation complied with the French National law, implementing the European Communities Council Directive 86/609/EEC. All efforts were made to minimize the number of animals used and their suffering, and the study met the ethical standards..

Tb Recording

Animals were anesthetized intraperitoneally by a mixture of 0.1ml of Rompun 2% (Xylazine) and 0.8ml of Zoletil 20 completed in a final volume of 5 ml with the NaCl 0.9 % (0.15ml/10g bodyweight). Anesthesia lasted more than one hour when implanting the miniature data loggers. Just before the injection of anesthetics we made the animal sniffing at cotton soaked with halothane (Laboratoires BELAMONT, Paris, FRANCE). A horizontal incision of less than 20 mm was made through the abdominal surface beforehand shaved in scissors and disinfected with alcohol solution and the data loggers were implanted. The incision was sutured closed and the wound sprayed with an antiseptic spray. After surgery, hypothermia was avoided with exposure in a heat coming from a light source. Generally, all animals recovered completely from implant surgery and their behavior it's normal.

The temperature-sensitive data loggers (DS1922L Thermochron iButtons®, Dallas Semiconductor, Texas, USA) used were coated by biocompatible silicone. The loggers were set to record the animals' Tb every 30 min with a resolution of 0.5°C and during 25 weeks. The data stored in i-buttons are got back at the end of the experiment via a reader connected directly with a computer by a bearing COM.

Testosterone Administration and Castration

Testosterone capsules (4-androsten-17β-ol-3-one, Sigma, Saint-Louis, MO, USA) measuring 10 mm were prepared with silastic tubes (DOW Coming, Midland, MI, USA; 1.57 mm internal diameter and 2.4 mm external diameter). Their extremities were occluded with silastic glue (DOW Corning). Testosterone capsules were stored at 4°C until use. In the Siberian hamster, Phodopus sungorus, a rodent of the same size as our animal model, capsules of this size filled with testosterone reproduce the physiological plasmatic concentrations [53]. The subcutaneous implantation of capsules was made by making a small latero-dorsal incision in the same time that of implantation of data loggers. Castrations were also performed by an abdominal incision in the same time that of implantation of data loggers.

Experimental Design

To establish the daily pattern of Tb rhythm in Lemniscomys barbarus, the effects of photoperiod and sexual steroids on daily rhythm of Tb, a total of adult's 20 males, trapped between July and October, 2009, were implanted with data loggers at the first week of December, 2009. The animals were subdivided into various experimental groups for Tb continuous recording during 25 weeks until May, 2010:

- Males animals were maintained in LP: 16/8 conditions (♂ LP group; n=5);
- Males animals were maintained in SP: 8/16 conditions, lights from 08:00 to 16:00 h (♂ SP group; n=5);
- Males animals were gonadectomised and transferred to SP: 8/16 conditions (♂ SP GDX group; n=5);
- Males animals were gonadectomised, implanted with testosterone capsules and transferred to SP: 8/16 conditions (♂ SP GDX-T group; n=5).

Data Analysis

All data are expressed as means ± S.E.M. Tb measurements were averaged for each half hour. Circadian rhythms were assessed by cosinor analysis using a nonlinear regression model. The analysis generated the following rhythm descriptors: mean level, amplitude and acrophase of Tb with SigmaPlot 11 software (Jandel Scientific, Chicago II, USA). To take into account the bimodal pattern of body temperature in Lemniscomys barbarus, a much better fitting was obtained with double-peak cosinor regression [6; 5] as follows:

$$f=y0+((a/((1+\exp(slope1*(t1-d-x)))*(1+\exp(slope1*(-t1-d+x)))))+(a2/((1+\exp(slope2*(t2-d-x)))*(1+\exp(slope2*(-t2-d+x))))))$$

Where y0 is the mean level, a the amplitude, t acrophase, d half-duration to reach the peak, a2 and t2 are the secondary amplitude and acrophase of double peak. We explored changes in Tb using ANOVA; Two Way Analysis of Variance was performed with using of 'Bonferroni t-test' for all pairwise multiple comparison procedures. The effects are considered as significant if $p<0.05$.

3 Results

3.1 Daily patterns of Tb: photoperiodic effects

The records of body temperature of a representative *Lemniscomys barbarus* maintained under a natural photoperiod are shown in figure (1-a). Clair daily rhythmicity is evident. Temperature oscillates between 36.41°C and 39.91°C, corresponding to normothermia, with a mean of 37.83 °C. To facilitate visual inspection of the synchronization between the body temperature rhythm and the light-dark cycle, Figure (1-b) and (1-c) show the grand means of temperature calculated by first averaging the body temperature readings at each time point over the 10 days for each animal and then averaging the data for the animals.

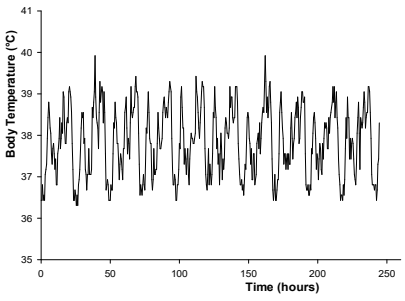
The majority of the highest values of daily thermal activation were positioned during the light-Phase. The cosinor regression of the daily profiles of T_b confirmed that the values variations are synchronized by the light-dark cycle (LD) as well in LP (Fig.1) as in SP conditions (Fig.2): the values of T_b rose during the light-phase and their reduction coincided widely with the night-phase.

Profiles of the average daily T_b rhythm corresponding to males exposed to LP (Fig.1) or SP (Fig.2) conditions show that the T_b rhythm is biphasic independently of photoperiodic regime. In LP conditions, there was a morning peak at 4.07±0.24 (Zeitgeber time) and later ones at 15.88±0.15. Peaks occurred during the light-phase which extends from 04:00-20:00h.

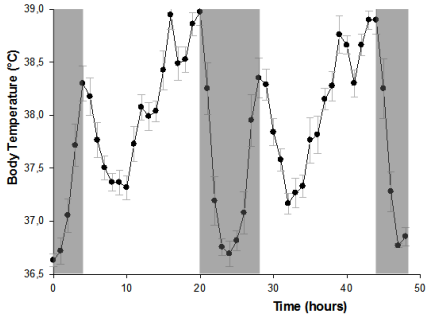
In SP conditions, the first acrophase is at 2.46±0.48 of dark-phase and the last one at 15.83±0.13 in the end of light- phase. In SP conditions, lights from 08:00 to 16:00h show that the first acrophase took place during the dark-period while the second one occurred just before light-offset.

The average T_b values in SP conditions was significantly higher in males in comparison to LP conditions (37.17 *vs.* 36.73, $p<0.001$; Bonferroni t-test). Nevertheless, the average of the first and second amplitude in LP were significantly higher than SP conditions (1.28 *vs.* 0.70, $p<0.001$; 3.17 *vs.* 2.26, $p<0.001$).

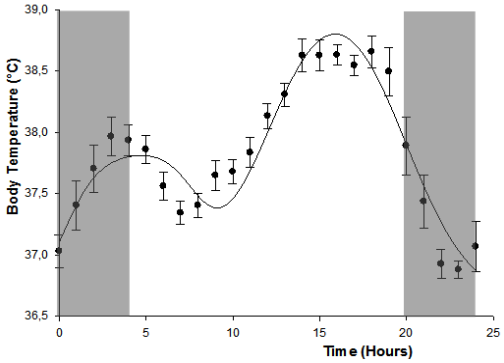
The acrophase was significantly differences between LP *vs.* SP for the first and second peak, respectively, 5.19 vs. 3.57 ($p<0.001$) and 15.08 *vs.* 15.02 ($p<0.001$). The first and second acrophase were significantly phase advanced in SP conditions compared to LP. Acrophase was phase advanced in SP than LP conditions.



(a)



(b)



(c)

Figure 1. Body temperature (Tb, °C) of Lemniscomys barbarus males during exposure to a long photoperiod (LD 16:8). (a): continuous measurement of Tb over 10 consecutive days, (b): mean (± SEM) body temperature, each averaged over 10 days, as function of time of day. (c): mean of Tb with cosinor regressions on 10 consecutive days of recording of Tb (—: express the best-fit non linear regressions). The grey shading indicates the dark phase.

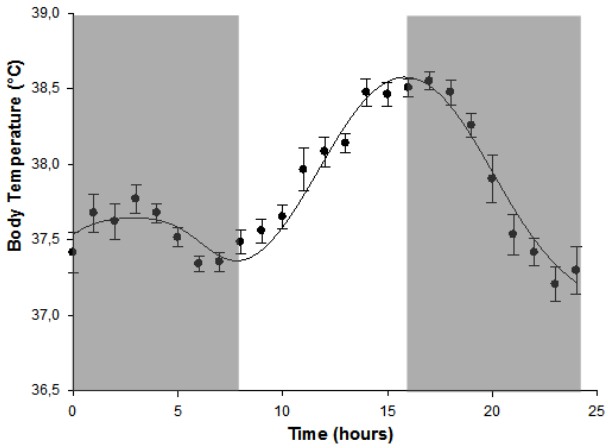


Figure 2. Daily profiles of the average of Tb rhythm in males Lemniscomys barbarus during exposure to a winter-like short photoperiod (LD 8:16). (—: express the best-fit non linear regressions). The grey shading indicates the dark phase.

3.2 Daily patterns of Tb: photoperiodic effects

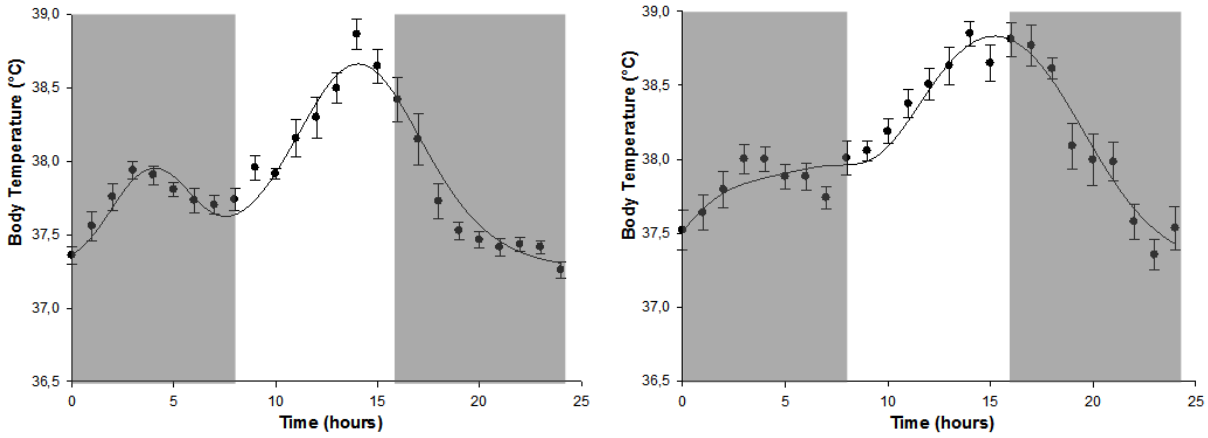
In SP conditions, castrated animals showed a biphasic daily T_b rhythm independently of testosterone administration or not (Fig. 3). In SP GDX group, the average value of T_b on the 10 consecutive analyzed days of measurement was 37.26±0.05°C, the first acrophase was at 3.92±0.22 and the second one was at 14.05±0.13. While the average for the first amplitude was 0.86±0.15, the second one was 2.73±0.39.

In SP GDX-T group, the average value of T_b on the 10 analyzed days of measurement was 37.25°C±0.12, the first acrophase was at 4.34±0.48 and the second one was at 15.20±0.19. The average for the first amplitude was 0.58±0.11 and 2.00±0.12 for the second one.

The comparison of GDX group *vs.* control group (intact males) shows that the average value of T_b was significantly higher in GDX group ($p<0.001$, 37.04 *vs.* 36.78; Bonferroni t-test). Also, the average of the first and second amplitude in GDX group were significantly higher than control group (1.16 vs. 0.94, $p<0.001$; 3.18 *vs.* 2.50, $p<0.001$). The first acrophase was significantly delayed in GDX group than control

group (4.72 *vs.* 3.26, p<0.001), however, the mean value of second acrophase was significantly advanced in GDX group than control group (14.07 *vs.* 15.85, p<0.001).

The average value of T_b was significantly higher in GDX-T group than control group (p<0.001, 37.03 *vs.* 36.78; Bonferroni t-test). The mean value of the first and second amplitude were significantly higher in control group than GDX-T one (0.94 *vs.* 0.87, p<0.001; 2.50 *vs.* 2.46, p<0.001). The first acrophase was significantly delayed in GDX-T group than control group (5.15 *vs.* 3.26, p<0.001). In the contrary, the second acrophase in GDX-T group was phase-advanced than in control group (15.22 *vs.* 15.85, p<0.001).



(a) (b)

Figure 3. Daily profiles of the average of Tb rhythm in males Lemniscomys barbarus under a short photoperiod condition (LD 8:16), castration (a) and testosterone administration (b). GDX and SP GDX-T. GDX, castrated animals and T, testosterone. The grey shading indicates the dark phase.

3.3 Daily torpor phenomenon occurrence

To determine if *Lemniscomys barbarus* engage in daily torpor phenomenon, data of continuous T_b recording in males exposed to SP: 8/16 conditions were minutely analysed between the 8th and the 25th weeks. In the present experimental conditions, no evidence of torpor bouts occurrence was observed. No sign of hibernation bouts was either observed, even in castrated animals. The T_b variations do not reach the characteristic values of daily torpor in small rodents (15°-20°C; [44]).

4 Discussion

By the use of miniature temperature-sensitive data loggers we were able to record continuously the T_b over several months. We confirm the diurnal feature of *Lemniscomys barbarus* already observed with wheel-running activity parameter [21]. We observed a strong rhythm of T_b in both photoperiodic conditions tested (LP and SP) in males. No arrhythmic individuals within our species can be reported, T_b was always rhythmic and diurnal under our laboratory conditions. A daily peak of T_b is observed during the light-phase that is logical for diurnal. Body temperature and locomotor activity rhythms are known to be robustly associated in time [38; 49]. Thanks to the determination of the profile of the daily T_b rhythm we were able to confirm the bimodal character of the daily rhythmicity in *Lemniscomys barbarus*, a character also already demonstrated in report to its wheel-running activity [21]. We know, indeed, that the bimodality is very close to the locomotor activity and the corticosterone rhythms [36; 52]. In our

animal model the daily T_b rhythm was biphasic independently of photoperiodic regime, *i.e.* LP or SP. Moreover, the biphasic behavior seems to be a genetically strong character in *Lemniscomys barbarus* since it doesn't alter by the gonads status.

Many authors explain the bimodal character in terms of ecophysiology by moving forward that this mode allows the animal to avoid the important hot temperatures of the middle-day. Similar bimodal activity was obtained in other diurnal species, i.e. *Arvicanthis ansorgei* [6], *Arvicanthis niloticus* [50] and *Octodon degus* [31; 37]. However, in other diurnal species as *Spermophilus tridecemlineatus*, *Canis familiaris*, *Equus caballus* and humans [40; 24; 34], the rhythm of T_b is unimodal even if the minimal values are registered during dark-phase as for *L. barbarus*.

This work demonstrates that daily T_b pattern is synchronized by the light-dark cycle and that *L. barbarus* shows a flexibility of this rhythm towards photoperiod variations; this result confirms the photoperiodic character of this species [21]. A study of *c-fos* expression rhythm shows indeed that the adjustment of the functioning of the endogenous clock is fast in long compared to short days [48; 54]. It would be interesting to verify, in our animal model, the profile of the daily rhythm of the core body temperature after a bilateral destruction of the SCN to confirm the role of circadian pacemaker of these nuclei. Furthermore, many studies were interested on the seasonality of the T_b rhythm [57; 1; 58] and the influence of photoperiod *via* its hormone the melatonin on the mechanisms of regulation of T_b was demonstrated; for example, in *Acomys russatus*, another diurnal species in whom the percentage of the T_b rhythmicity is lesser in SP than in LP and the exogenous melatonin has a repressor effect on this rhythm [16; 17].

In this work we show that the T_b values oscillate between 36.41°-39.91°C in males whatever the photoperiodic regime. These changes are almost similar to others endotherms who display a daily oscillation in T_b [26]. This relatively large range of T_b rhythm oscillations may have beneficial effect on thermoregulation in an environment with daily ambient temperature cycles like at Tétouan latitude (North of Morocco). Nevertheless, the range of T_b oscillation obtained in *Lemniscomys barbarus* was slightly higher compared with that of a dark-active rodent like the Syrian hamster, *e.g.*, 2.9°C [39]. Barbary striped grass mouse is a very lively and very active rodent as demonstrated by Lahmam (2008) [21] thanks to the wheel-running activity and the accentuate range of T_b oscillations can be due to an increased activity as confirmed in several species [38]. Besides, in the Djungarian hamster, *Phodopus sungorus*, a similar-sized rodent than our model animal, mean T_b was 33.18°C±1.4 in the warm (20.0°C±0.86) [51].

In our diurnal rodent, *Lemniscomys barbarus*, photoperiodic effects on the daily T_b rhythm were observed. The amplitude of the daily T_b rhythm was higher in LP than in SP conditions. In the Siberian hamster, a similar-sized rodent, several daily rhythms were also lower in amplitude in short days than in long ones [35].

Concerning the sexual steroids action we observed that testosterone affects the acrophase timing (phase advance peak), the amplitude and the average of the T_b rhythm. In both group, GDX and GDX-T the mean values of T_b were higher than control group. Indeed, the first acrophase was significantly advanced in control group than GDX and GDX-T groups. However, the second acrophase was phase-delayed in control group than GDX and GDX-T groups. The effects obtained after experimental processes such as the castration or the exogenous testosterone administration, which remain artificial approaches, are contrasted by the fact that in non treated animals exposed to LP or SP conditions the spontaneous testicular activation and gonads atrophy, respectively, can be of a variable level.

It was indeed reported that the testosterone plays a role especially in the organization of the daily T_b pattern [29]. Because sexual steroids concentrations in LP are supposed constant even after 25 weeks, average T_b, the amplitude and the acrophase did not change according to time in LP. Besides, since the age of animals captured in the nature is difficult to estimate, we cannot exclude the effect of the age which can also explain the characteristics changes of the T_b rhythm. The animal's age, indeed, can alter the circadian rhythm of T_b [28; 1; 56].

To survive in the rigorous conditions of the bad season, small-sized heterothermic mammals develop various strategies of seasonal acclimatization among which the expression of the daily torpor [55; 14]. This annual function is essentially controlled by photoperiod and ten weeks of SP exposure in laboratory conditions are enough for leading the torpor appearance [33]. In Tétouan region, the captures of *Lemniscomys barbarus* being much less successful during the bad season than during the spring and summer, we strongly suspected this rodent of living in lethargy; the torpor bouts expression was then investigated.

In our experimental conditions with food *ad libitum* and thermoneutrality, *L. barbarus* did not express episodes of torpor further to a long exposure in SP conditions (SP: 8/16) in intact and castrated animals. In the other hand, the mouse *Peromyscus maniculatus* is sexually inactive when the food becomes rare [3]. So, the daily torpor phenomenon in certain species of mammals can be led by a food deprivation [47]. In this work, we tried to highlight daily torpor while animals were in condition of food in excess. It is possible that in *Lemniscomys barbarus* a partial hardship of the food could potentiate the effect of torpor induction by SP. It's well know indeed that in temperate regions heterothermic regulations are most pronounced during long and severe winters when low ambient temperatures are normally associated with reduced food availability [13]. In future works, it would be of a certain interest to verify this track in *Lemniscomys barbarus* maintained in SP conditions associated to cold and food restriction.

5 Conclusion

This study showed for the first time the profile of daily T_b rhythm in *Lemniscomys barbarus*, characterized by a strong rhythm in both photoperiodic conditions (LP and SP) with a daily peak during the light phase which confirms the diurnal feature of our species.

According to the experimental conditions of this work, we were not able to establish if *Lemniscomys barbarus* engage on the phenomenon of daily torpor. Nevertheless, this characteristic mechanism of energy saving in the small-sized mammals could be demonstrated in a future study by the combined effects of SP, low ambient temperature and reduction of food access.

ACKNOWLEDGMENT

The thanks of the authors are sent to Pr. Mohammed Errami, Tétouan, Morocco, for all the efforts he displayed to provide the equipment required for the animal facilities. The authors sent deeply thanks for their valuable contribution to Dr. Ibtissam Chakir from our laboratory of biology and health, faculty of sciences, Tétouan, and Dr. Patrick Vuillez from the Institute for Integrative Cellular and Neurosciences, UPR 3212 University of Strasbourg CNRS, Strasbourg, France.

REFERENCES

[1] Benstaali C., Bogdan A., Touitou Y., 2002. Effect of a short photoperiod on circadian rhythms of body temperature and motor activity in old rats. *Pflugers Arch*. **444**(1-2):73-79.

[2] Biggar K.K., Wu C.W., Tessier S.N., Zhang J., Pifferi F., Perret M., Storey K.B., 2015. Primate Torpor: Regulation of Stress-activated Protein Kinases During Daily Torpor in the Gray Mouse Lemur, *Microcebus murinus*. *Genomics Proteomics Bioinformatics*. **13**:81-90.

[3] Blank J.L., Desjardins C., 1984. Spermatogenesis is modified by food intake in mice. *Biol. Reprod.* **30**: 410-415.

[4] Challet E., Pitrosky B., Sicard B., Malan A., Pévet P., 2002. Circadian organization in a diurnal rodent, *Arvicanthis ansorgei* Thomas 1910: chronotypes, responses to constant lighting conditions and photoperiodic changes. *J. Biol. Rhythms*. **17**:52–64.

[5] Chakir I., 2014. Rôle du gène horloge Clock dans l'intégration de la photopériode par les noyaux suprachiasmatiques chez un rongeur nocturne, Mesocricetus auratus, et dans le déterminisme de la diurnalité chez un rongeur diurne, *Lemniscomys barbarus*. Doctoral thesis at the Abdelmalek Essaâdi University, Tétouan, Morocco.

[6] Cuesta M., Clesse D., Pévet P., Challet E., 2009. From daily behavior to hormonal and neurotransmitters rhythms: Comparison between diurnal and nocturnal rat species. *Hormones and Behavior*. **55**: 338-347.

[7] Dardente H., Menet J.S., Challet E., Tournier B.B., Pévet P., Masson-Pévet M., 2004. Daily and circadian expression of neuropeptides in the suprachiasmatic nuclei of nocturnal and diurnal rodents. *Mol. Brain. Res*. **124**:143–151.

[8] Dzal Y.A., Brigham R.M., 2013. The tradeoff between torpor use and reproduction in little brown bats (Myotis lucifugus). *J. Comp. Physiol.B*. **183** : 279-288.

[9] El Allali K., Achaâban M.R., Bothorel B., Piro M., Bouâouda H., El Allouchi M., Ouassat M., Malan A., Pévet P., 2013. Entrainment of the circadian clock by daily ambient temperature cycles in the camel (*Camelus dromedarius*). *Am J Physiol Regul Integr Comp Physiol*; **304**(11):R1044-52.

[10] El Ouazzani S., Janati I.A., Magoul R., Pévet P., Saboureau M., 2011. Overwinter body temperature patterns in captive jerboas (*Jaculis orientalis*): Influence of sex and group. *J. Comp. Physiol. B*. **181**(2):299-309.

[11] Freeman M.E., Crissma J.K., Low G.N., Butcher R.L., Inskeep E.K., 1970. Thermogenic action of progesterone in the rat. *Endocrinology*. **86**: 717-720.

[12] Geiser F., 2004. Metabolic rate and body temperature reduction during hibernation and daily torpor. Annu. Rev. Physiol. **66**: 239-274.

[13] Geiser F., Ruf T., 1995. Hibernation *versus* daily torpor in mammals and birds: physiological variables and classification of torpor patterns. *Physiol. Zool*. **68**:935–66.

[14] Geiser F., Turbill C., 2009. Hibernation and daily torpor minimize mammalian extinctions. *Naturwissenschaften* .**96**:1235-1240.

[15] Golombek D.A., Rosenstein R.E., 2010. Physiology of circadian entrainment. *Physiol. Rev*. **90**:1063-1102.

[16] Haim A., Zizapel N., 1994. Daily rhythm of body temperature in coexisting rodents of the *Genus Acomys* acclimated to long-photoperiod: effects of ethanol and melatonin. *Thermal Balance in Health and Disease APS*. **22**:191-196.

[17] Haim A., Shachaf K., Zisapel N., Reiter R.J., 1997. Daily rhythm of body temperature in *Acomys russatus*: Their response to photoperiod manipulations and melatonin. *J. therm. Biol*. **22**(3): 219-222.

[18] Heldmaier G., Steinlechner S., 1981. Seasonal pattern and energetic of short daily torpor in the Djungarian hamster, *Phodopus sungorus*. *Oecologia* .**48** : 265-270.

[19] Jefimow M., Ostrowski M., Jakubowska A., Wojciechowski M.S., 2014. The effects of Dietary Cholesterol on metabolism and Daily Torpor Patterns in Siberian Hamsters. *Physiological and Biochemical Zoology*. **87**(4): 527-538.

[20] Kirsch R., Ouarour A., Pévet P., 1991. Daily torpor in the Djungarian hamster (*Phodopus sungorus*): photoperiodic regulation, characteristics and circadian organization. *J. Comp. Physiol. A*. **168**:121-128.

[21] Lahmam M., El M'rabet A., Ouarour A., Pévet P., Challet E., Vuillez P., 2008. Daily behavioral rhythmicity and organization of the suprachiasmatic nuclei in the diurnal rodent, *Lemniscomys barbarus*. *Chronobiol Int*. **25** :882-904.

[22] Lahmam M., 2008. Rythmicité comportementale journalière et organisation des noyaux suprachiasmatiques et de la rétine chez *Lemniscomys barbarus*, un rongeur diurne à reproduction saisonnière. Doctoral Thesis at the Abdelmalek Essaâdi University, Tétouan, Morocco.

[23] Lee T.M., Labyak S.E., 1997. Free-running rhythms and light- and dark-pulse phase response curves for diurnal *Octodon degus* (Rodentia). *Am. J. Physiol. Regul. Integr. Comp. Physiol*. **273**: R278–R286.

[24] Leproult R., Van Onderbergen A., L'hermite-Baleriaux M., Van Cauter E., Copinschi G., 2005. Phase-Shifts of 24-h rhythms of hormonal release and body temperature following early evening administration of melatonin agonist agomelatine in healthy older men. *Clin, Endocrinol*. (Oxf) .**63**:298-304.

[25] Levi O., Dayan T., Kronfeld-Schor N., 2011. Adaptive thermoregulation in Golden Spiny Mice: The Influence of Season and Food Availability on Body Temperature. *Physiological an Biochemical Zoology*. **84**(2): 175-184.

[26] Lovegrove B.G., Heldmaier G., Ruf T., 1991. Perspectives of endothermy revisited: the endothermic temperature range. *J Therm Biol* .**16**(4): 185–197.

[27] Mahoney M., Bult A., Smale L., 2001. Phase response curve and light-induced fos expression in the suprachiasmatic nucleus and adjacent hypothalamus of *Arvicanthis niloticus*. *J. Biol. Rhythms* .**16**: 149–162.

[28] Maillox A., Benstaali C., Bogdan A., Auzéby A., Touitou Y., 1999. Body temperature and locomotor activity as marker rhythms of aging of the circadian system in rodents. *Experimental Gerontology*. **34**(6): 733-740.

[29] Marrone B.L., Gentry R.T., Wade G.N., 1976. Gonadal Hormones and Body Temperature in Rats: Effects of Estrous Cycles, Castration and Steroid Replacement. *Physiology & Behavior*. **17**(3): 419-425.

[30] McDonald R.B., Day C., Carlson K., Stern J.S., Horwitz B.A. 1989. Effect of age and gender on thermoregulation. *Am. J. Physiol*. **257** (4 pt 2), R700-R704.

[31] McElhinny T.L., Smale L., Holeklamp K.E., 1997. Patterns of body temperature, activity, and reproductive behavior in a tropical murid rodent, *Arvicanthis nioticus*. *Physiol. Behav*. **62**: 91-96.

[32] Mzilikazi, N., Lovegrove, B.G., 2002. Reproductive activity influences thermoregulation and torpor in the pouched mouse, *Saccostomus campestris*. *J. comp. Physiol. B*. **172**: 7-16.

[33] Ouarour A., Kirsch R., Pévet P., 1991. Effects of temperature, steroids and castration on daily torpor in the Djungarian hamster (*Phodopus sungorus*). *J. Comp. Physiol. A*. **168**:477-481.

[34] Piccione G., Caola G., Refinetti R., 2005. Temporal relationships of 21 physiological variable in horse and sheep. *Com. Biochem. Physiol. A. Mol. Integr. Physiol*. **142**:389-396.

[35] Prendergast B.J., Tyler J. Stevenson T.J., Zucker I., 2013. Sex differences in Siberian hamster ultradian locomotor rhythms. *Physiol. Behav.* 0: 206–212.

[36] Refinetti R., Manaker, M., 1992. The circadian rhythm of body temperature. *Physiol.Behav*. **51**:613-637.

[37] Refinetti R., 1996. Rhythm of body temperature and temperature selection are out of phase in a diurnal rodent, *Octodon degus*. *Physiol. Behav*. **60**:959-961.

[38] Refinetti R. 1999a. Relationship between the daily rhythms of locomotor activity and body temperature in eight mammalian species. *Am. J. Physiol*. **277** (5pt2): R1493–R1500.

[39] Refinetti R., 1999b. Amplitude of the daily rhythm of body temperature in eleven mammalian species. *J Thermal Biol*. **24** (5-6): 477–481.

[40] Refinetti R., Piccione, G., 2004. Intra-and inter-individual variability in the circadian rhythm of body temperature of rats, squirrels, dogs, and horses. *J. of Thermal Biology*. **30**:139-146.

[41] Refinetti R., 2006. Variability of diurnality in laboratory rodents. *J. Comp. Physiol. A*. **192**(7):701-714.

[42] Refinetti, R., 2010. The circadian rhythm of body temperature. *Frontiers in Bioscience*. **15**: 564–594.

[43] Ruf T., Geiser F., 2015. Daily torpor and hibernation in birds and mammals. *Biol. Rev*. **90**:891-926.

[44] Ruf T., Heldmaier G., 1992. The impact of daily torpor on energy requirements in the Djungarian hamster, *Phodopus sungorus*. *Physiol Zool* **65**: 994–1010.

[45] Schmidt P.T., Raju, J., Danaceau M., Murphy D.L., Berlin R.E., 2002. The effects of gender and gonadal steroids on the neuroendocrine ant temperature response to m-chlorophenylpiperazine in leuropide-induced hypogonadism in women and men. *Neuropsychopharmacology*. 27800-27812.

[46] Smale L., Lee T., Nunez A.A., 2003. Mammalian diurnality: some facts and gaps. *J. Biol. Rhythms*. **18**:356–366.

[47] Steinlechner S., Heldmaier G., Weber C., Ruf T., 1986. Role of photoperiod: pineal gland interaction in torpor control. In Heller H.C. *et al.* (Eds), Living in the cold: 301-307.

[48] Sumova A., Travnickova Z., Illnerova H., 1995. Memory on long but not on short days is stored in the rat suprachiasmatic nucleus. *Neurosci. Lett*. **200**(3): 191-4.

[49] Tachinardi P., Bicudo J.E.W., Oda G.A., Valentinuzzi V.S, 2014. Rhythmic 24 h Variation of Core Body Temperature and Locomotor Activity in a Subterranean Rodent (*Ctenomys* aff. *knighti*), the Tuco-Tuco. PLoS ONE **9**(1): e85674.

[50] Teresa L., McElhinny, Smale, L., Holekamp, K.E., 1997. Patterns of body temperature, Activity, and reproductive behavior in a tropical murid roden, *Arvivanthis niloticus*. *Physiology & Behavior*. **62**(1): 91-96.

[51] Turbill C., Smith S., Deimel C., Ruf T., 2012. Daily torpor is associated with telomere length change over winter in Djungarian hamsters. *Biol. Lett.* **8**:304–307.

[52] Verhagen L.A., Pévet P., Saboureau M., Sicard B., Nesme B., Claustrat B., Buijs R.M., Kalsbeek A., 2004. Temporal organization of the 24-h corticosterone rhythm in the diurnal murid rodent *Arvicanthis ansorgei* Thomas 1910. *Brain Res*. **995**(2):197-204.

[53] Vitale P.M., Darrow J.M., Duncan M.J., Shustak C.A., Goldman B.D., 1985. Effects of photoperiod, pinealectomy and castration on body weight and daily torpor in Djungarian hamsters (*Phodopus sungorus*). *J. Endocrinology*. **106**:367-375.

[54] Vuillez P., Jacob N., Teclemariam-Mesbah R., Pévet P., 1996. In Syrian and European hamsters, the duration of sensitive phase to light of the suprachiasmatic nuclei depends on the photoperiod. *Neurosci Lett*.**208**(1):37-40.

[55] Wang L.C.H., Wolomyk M.W., 1988. Torpor in mammals and birds. *Can. J. Zool*. **66**: 133-137.

[56] Weinert D., 2010. Circadian temperature variation and ageing. *Ageing Res. Rev*. **9**(1): 51-60.

[57] Wollnik, F., Schimidt, B., 1995. Seasonal and rhytms of body temperature in the European hamster (*Cricetus cricetus*) under semi-natural conditions. *J. Comp Physiol B*. **165**: 171-182

[58] Zhi-Jun Zhaon, Jing Cao, Zheng-Chao Liu, Gui-Ying Wang, Lu-Sheng Li, 2010. Seasonal regulations of resting metabolic rate and thermogenesis instriped hamster (*Cricetulus barabensis*). *J. of Thermal Biology*. **35**: 401-405.

[59] Zuloaga D.G., McGiverna R.F., Handa R.J., 2009. Organizational influence of the postnatal testosterone surge on the circadian rhythm of core body temperature of adult male rats. *Brain Research*. **1268**:68-75.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Fuzzy Robust $H\infty$ Tracking Control For Wind Generator System: LMI approach

[1]Kaoutar Lahmadi, [2]Ismail Boumhidi
[1,2]Department of physics, LESSI Laboratory, Faculty of Sciences Dhar el Mahraz, University Sidi Mohammed Ben Abdellah, Fez, Morocco
kaoutar.lahmadi@usmba.ac.ma, iboumhidi@hotmail.com

ABSTRACT

This study concerns the tracking control problem of the wind turbine generator system with uncertainties parameters and external disturbances. Based on T-S fuzzy model, a fuzzy observer-based and a fuzzy robust state feedback output tracking control are developed to reduce the tracking error by minimizing the disturbance level caused by the wind speed. Using a Lyapunov function combined with H∞ tracking criteria and a judicious of the famous Young relation, a sufficient stability condition for the robust fuzzy tracking control formulated in terms of linear matrix inequality, which can be very efficiently solved by using LMI optimization techniques. The simulation results are given to show the performance of the observer-based tracking controller.

Keywords—Wind system, H∞ tracking control, observer-based controller, linear matrix inequality (LMI).

1 Introduction

The Takagi –Sugeno (T-S) fuzzy model approach has nowadays become popular since [1][2][6], it showed its efficiency to control complex nonlinear systems with uncertainties parameters and used for many applications [7]. This fuzzy dynamic model is a piecewise interpolation of several linear models through membership functions [4]. It is described by fuzzy rules of the type IF-THEN that represent local input output models for a nonlinear system [2]. For more details on these topics we refer the reader to Tanaka and Wang [6].

For a few years, the trend of fuzzy tracking control for uncertain nonlinear systems has attracted great attention [14] [8]. The main issue in the control theory is that of controlling a system in order to have its output asymptotically tracking reference signals [10]. Despite an abundant literature on stability conditions of T-S models, few authors have dealt with the tracking problem. Some works are concerned with state feedback, H∞ performances and output feedback [12]. The most important issue for fuzzy tracking control systems is how to reduce the tracking error between the desired trajectory and the actual output values rapidly with the guaranteed stability [16]. In this work a sufficient condition are derived for stabilization of the robust fuzzy tracking controller and the robust fuzzy observer in the sense of Lyapunov asymptotic stability [15].

DOI: 10.14738/tmlai.54.2971
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2971

In the last decade, optimal H∞ control theory has been well developed and found extensive application to efficiently treat the robust stabilization and disturbance rejection problems. The H∞ control performance for uncertain nonlinear systems is proposed to attenuate the effects caused by modeled dynamics, disturbances and approximate errors [13].

The task of tracking is a typical control problem in the industry. In this study we focused in control of wind system formed by a turbine, a driving shaft and a double fed induction machine .As well known, the primary energy source of the wind generators is the wind which is unpredictable and varies frequently [9]. Wind energy has greatly progressed by becoming one of most renewable energy competitive [5].The control plays a very important role to make wind technology more profitable and reliable. Then an H∞ output feedback controller is designed by taking into account directly wind speed as perturbation [15].

In this paper, the tracking problem for wind turbine is studied using the T-S fuzzy approach. First a fuzzy augmented system is constructed by regrouping the T-S fuzzy model, a fuzzy observer, a fuzzy tracking control. Then, based on Lyapunov function combined with a Young Relation and H∞ criteria a new condition for stabilization of the whole system is obtained. This condition is applied on wind turbine system with uncertainties parameters and external disturbances in order to guarantying the performances and to reduce the tracking error by minimizing the disturbance level caused by the wind speed. The proposed methodology is formulated as the linear matrix inequality problem (LMIs) which can be solved very efficiently using the convex optimization techniques.

The paper is arranged as follows: the uncertain T-S fuzzy model with external disturbances, the tracking criteria as well as the observer design and the tracking control law are described in section (II). In section (III) a sufficient condition based on the observer-based tracking control and using Lyapunov function is developed in order to in to minimize the tracking error. Section (IV) presents the performances of the proposed method according to robustness tests are shown through simulations of the wind turbine in. Section (V) concludes this paper.

2 Problem formulation

2.1 T-S fuzzy model

Let consider the T-S fuzzy system with uncertainties parameters. The i-th rule is described as follows:

Plant rule:

If $z_1(t)$ is N_{ij} and … and $z_p(t)$ is N_{ip}

Then
$$\begin{cases} \dot{x}(t) = (A_i + \Delta A_i)x(t) + (B_i + \Delta B_i)u(t) + B_{2i}\varphi(t) \\ y(t) = C_i x(t) \end{cases} \quad (1)$$

Where N_{ij} a fuzzy set is, $x(t) \in R^n$ is the system state vector, $u(t) \in R^m$ is the control input vector, $\varphi(t) \in R^p$ is the disturbance input vector, $y(t) \in R^p$ is the system output, A_i, B_i, C_i are known constant matrix that describe the nominal system, $z_1(t) = [z_1(t), z_2(t), \ldots\ldots, z_p(t)]$ are the premise variables. The Lebesgue measurable uncertainties are defined as

$\Delta A_i(t) = H_i F_i(t) E_{ai}$, $\Delta B_i(t) = H_i F_i(t) E_{bi}$, where matrices H_i, E_{ai} and E_{bi} are constant of appropriate dimension and $F(t)$ is unknown matrix function which is bounded by: $F_i^T(t) F_i(t) \le I$

Given a pair of $(x(t), u(t))$, the final outputs of the fuzzy systems are inferred as follows:

$$\dot{x}(t) = \frac{\sum_{i=1}^{r} \mu_i(z(t)) \left[(A_i + \Delta A_i(t)) x(t) + (B_i + \Delta B_i(t)) u(t) + B_{2i} \omega(t) \right]}{\sum_{i=1}^{r} \mu_i(z(t))}$$

Where $\mu_i = \prod_{i=1}^{n} N_{ij} z(t)$ and $h_i(z(t)) = \frac{\mu_i(z(t))}{\sum_{i=1}^{r} \mu_i(z(t))}$

The defuzzification process of the TS fuzzy model (1) with uncertainties parameters can be represented as:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^{r} h_i(z(t)) \left[(A_i + \Delta A_i(t)) x(t) + (B_i + \Delta B_i(t)) u(t) + B_{2i} \varphi(t) \right] \\ y(t) = \sum_{i=1}^{r} h_i(z(t)) C_i x(t) \end{cases} \quad (2)$$

$i = 1, 2 \ldots\ldots, \; r$

2.2 Observer design

In the present work, the case where the state variables are not available for the feedback control with unmeasurable premise variable is considered.Therfore, the following T-S observer is considered to deal with the state estimation of T-S nonlinear system (2).

$$\begin{cases} \dot{\hat{x}}(t) = \sum_{i=1}^{r} A_i \hat{x}(t) + B_i u(t) + G_i (y(t) - \hat{y}(t)) \\ \hat{y}(t) = \sum_{i=1}^{r} C_i \hat{x}(t) \end{cases} \quad (3)$$

Where $G_i (i = 1, \ldots., \; r)$ are the observers gains to be determined and $\hat{x}(t)$ is the state estimation.

Let us define the state estimation error $e_x(t)$ as:

$$e_x(t) = x(t) - \hat{x}(t) \;\; \dot{e}_x(t) = \dot{x}(t) - \dot{\hat{x}}(t) \quad (4)$$

2.3 Tracking criteria

To deal with the tracking problem for the T-S fuzzy system, we consider the following reference model:

$$\dot{x}_r(t) = A_r(t) x_r(t) + r(t) \quad (5)$$

$x_r(t)$ the reference state;

$A_r(t)$ a specified asymptotically stable matrix;

$r(t)$ a bounded reference input.

Then the objective is to design a T–S fuzzy model-based controller, which stabilizes the fuzzy system (2) and achieves the H_∞ tracking performance related to tracking error $e_r(t)$ as follows [tong 2002]:

$$\int_0^{t_f} e_r^T(t) Q e_r(t) d(t) \le \eta^2 \int_0^{t_f} \bar{\phi}^T(t) \bar{\phi}(t) d(t) \quad (6)$$

Where t_f denotes the final time, Q is a positive definite weighting matrix, and η is a specified attenuation level.

The tracking error is defined by:

$$e_r(t) = x(t) - x_r(t)\ \dot{e}_r(t) = \dot{x}(t) - \dot{x}_r(t) \tag{7}$$

2.4 Tracking control:

Suppose the following fuzzy controller is employed to deal with the above control system design, the control structure is choosen as a PDC law:

$$u(t) = -\sum_{i=1}^{r} h_i(z(t))k_i(x_r(t) - \hat{x}(t)) \tag{8}$$

Where k_i are gain matrices with appropriate dimension.

Let us consider the estimation error $e_x(t) = x(t) - \hat{x}(t)$, the tracking error state reference $e_r(t) = x(t) - x_r(t)$ and the state reference $x_r(t)$ .By substituting (2),(3) and (8)in (4),and substituting (2), (5)and (8) in (7) we get :

$$\dot{e}_x(t) = \sum_{i=1}^{r} h_i(z(t))[(A_i - L_i C_i)e_x(t) + \Delta A(t)e_r(t)$$

$$+ \Delta A(t)x_r(t) + \Delta B(t)k_j(e_r(t) - e_x(t)) + B_{2i}\omega(t)]$$

$$\dot{e}_r(t) = \sum_{i=1}^{r} h_i(z(t))[(A_i - A_r + \Delta A(t))x_r(t) + (A_i + \Delta A(t))e_r(t)$$
$$+ (B_i + \Delta B(t))k_j(e_r(t) - e_x(t)) + B_{2i}\omega(t) - r(t)]$$

$$\dot{x}_r(t) = A_r(t)x_r(t) + r(t)$$

The state vector for the global closed loop is $x_a^T(t) = (e_r^T(t) \quad e_x^T(t) \quad x_r^T(t))$

Some easy manipulations lead to the following augmented system:

$$\dot{x}_a(t) = \sum_{i=1}^{r}\sum_{j=1}^{r} h_i(z(t))h_j(z(t))(\bar{A}_{ij}(t)x_a(t) + \bar{S}\bar{\phi}(t)) \tag{9}$$

Where:

$$\bar{\phi}(t) = \begin{bmatrix} r(t) \\ \varphi(t) \end{bmatrix} \qquad \bar{S}(t) = \begin{bmatrix} -I & B_{2i} \\ 0 & B_{2i} \\ I & 0 \end{bmatrix}$$

$$\bar{A}_{ij}(t) = \begin{bmatrix} A_i + \Delta A_i(t) + B_i k_j + \Delta B_i(t)k_j & -B_i k_j - \Delta B_i(t)k_j & A_i - A_r + \Delta A_i(t) \\ \Delta A_i(t) + \Delta B_i(t)k_j & A_i - L_j C_i - \Delta B_i(t)k_j & \Delta A_i(t) \\ 0 & 0 & A_r \end{bmatrix}$$

Hence, the tracking criteria H_∞ (6) with the augmented vector $x_a(t)$ can be modified as follows:

$$\int_0^{tf} x_a^T(t)Q_a x_a(t)d(t) \le \eta^2 \int_0^{tf} \bar{\phi}^T(t)\bar{\phi}(t)d(t) \tag{10}$$

With $Q_a = (Q,0,0,0)$ and $\bar{\phi}^T(t)\bar{\phi}(t) = r^T(t)r(t) + \varphi^T(t)\varphi(t)$

3 LMI Formulation of the Fuzzy Observer-Based Tracking Control

The design purpose in this study is to determine the gains k_j and L_j described in (9) to ensure the asymptotic stability of the closed loop system (9) and achieve the H_∞ tracking performance in (10).The main result on the fuzzy tracking control design of T-S fuzzy model with uncertainties parameters are summarized in the following theorem:

Theorem .Given the uncertain system described by (2).If there exist a symmetric and positive definite $P = P^T \succ 0$ and $Q \succ 0$ some matrices X_j, W_j and a positive scalars η , μ_1, μ_2, μ_3 and μ_4, such that the following condition (11) hold for $i, j = (1,\ldots., r)$:

$$\begin{bmatrix} \Sigma_{11} & N & -B_i X_j & \Sigma_{14} & N^T E_{ai}^T & N^T E_{ai}^T & X_j^T E_{bi}^T & -X_j^T E_{bi}^T & -N & B_{2i} N \\ * & -Q^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & * & \Sigma_{33} & 0 & 0 & 0 & 0 & 0 & 0 & P_2 B_{2i} \\ * & * & * & \Sigma_{44} & 0 & 0 & 0 & 0 & P_3 & 0 \\ * & * & * & * & -\mu_1^{-1} I_1 & 0 & 0 & 0 & 0 & 0 \\ * & * & * & * & * & -\mu_2^{-1} I_2 & 0 & 0 & 0 & 0 \\ * & * & * & * & * & * & -\mu_3^{-1} I_3 & 0 & 0 & 0 \\ * & * & * & * & * & * & * & -\mu_4^{-1} I_4 & 0 & 0 \\ * & * & * & * & * & * & * & * & -\eta^2 I_5 & 0 \\ * & * & * & * & * & * & * & * & * & -\eta^2 I_6 \end{bmatrix} \prec 0 \tag{11}$$

With:

$$\Sigma_{11} = N_i A_i^T + A_i N + X_j^T B_i^T + B_i X_j + (\mu_1^{-1} + \mu_2^{-1} + \mu_3^{-1} + \mu_4^{-1}) H_i H_i^T$$
$$\Sigma_{14} = A_i N - A_r N$$
$$\Sigma_{33} = A_i^T P_2 + P_2 A_i - W_j^T C_i^T - C_i W_j$$
$$\Sigma_{44} = A_r^T P_3 + P_3 A_r$$

The controller and observer gains are given $j = (1,\ldots,r)$ by $N = P_1^{-1}$, $X_j = k_j P_1^{-1}$, $W_j = P_2 L_j$

Proof. Let us now consider the following candidate Lyapunov function for the augmented system (9):

$$V(t) = x_a^T(t) P x_a(t) \tag{12}$$

Where the weighting matrix $P = P^T \succ 0$

Using (9) the time derivative of $V(t)$ is the following by

$$\dot{V}(t) = \dot{x}_a^T(t) P x_a(t) + x_a^T(t) P \dot{x}_a(t) \tag{13}$$

Substituting (9) into (13) we get:

$$\dot{V}(t) = \sum_{i=1}^{r} \sum_{j=1}^{r} h_i(z(t)) h_j(z(t)) [x_a^T(t)(\overline{A}_{ij}^T(t) P + P^T \overline{A}_{ij}(t)) x_a(t)$$
$$+ \overline{\phi}^T(t) \overline{S}^T P x_a(t) + x_a^T(t) P^T \overline{S} \overline{\phi}(t)] \tag{14}$$

To ensure the stability of the augmented system (9) and guarantying the H_∞ performances in (10), the condition to verify is:

$$\dot{V}(t) + x_a^T(t) Q_a x_a(t) - \eta^2 \overline{\phi}^T(t) \overline{\phi}(t) \prec 0 \tag{15}$$

From equations (14) and (9), (15) becomes: $\sum_{i=1}^{r} h_i(z(t)) \sum_{j=1}^{r} h_j(z(t)) \left[x_a^T(t) \quad \bar{\phi}^T(t) \right]$

$$\begin{bmatrix} \bar{A}_{ij}^T P + P\bar{A}_{ij} + Q_a & P\bar{S} \\ \bar{S}^T P & -\eta^2 I \end{bmatrix} \begin{bmatrix} x_a(t) \\ \bar{\phi}(t) \end{bmatrix} \prec 0 \tag{16}$$

This is satisfied if:

$$\begin{bmatrix} \bar{A}_{ij}^T P + P\bar{A}_{ij} + Q_a & P\bar{S} \\ \bar{S}^T P & -\eta^2 I \end{bmatrix} \prec 0 \tag{17}$$

Fortunately (16) can be transferred into a LMI problem which can be solved by using convex optimization techniques. Let us assume that $P = diag\ (P_1, P_2, P_3)$ with $P_1 = P_1^T \succ 0, P_2 = P_2^T \succ 0, P_3 = P_3^T \succ 0$

Substitution of P into (16) yields:

$$\begin{bmatrix} e_r^T(t) \\ e_x^T(t) \\ x_r^T(t) \\ r^T(t) \\ \varphi^T(t) \end{bmatrix}^T \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & -P_1 & P_1 B_{2i} \\ * & Y_{22} & Y_{23} & 0 & P_2 B_{2i} \\ * & * & Y_{33} & P_3 & 0 \\ * & * & * & -\eta^2 I & 0 \\ * & * & * & * & -\eta^2 I \end{bmatrix} \begin{bmatrix} e_r(t) \\ e_x(t) \\ x_r(t) \\ r(t) \\ \varphi(t) \end{bmatrix} \prec 0 \tag{18}$$

With:

$Y_{11} = A_i^T P_1 + P_1 A_i + k_j^T B_i^T P_1 + P_1 B_i k_j + \Delta A_i^T(t) P_1$

$+ P_1 \Delta A_i(t) + k_j^T \Delta B_i^T(t) P_1 + P_1 \Delta B_i(t) k_j + Q$

$Y_{12} = \Delta A_i^T(t) P_2 + k_j^T \Delta B_i^T(t) P_2 - P_1 \Delta B_i(t) k_j - P_1 B_i k_j$

$Y_{13} = P_1 A_i - P_1 A_r + P_1 \Delta A_i(t)$

$Y_{22} = A_i^T P_2 + P_2 A_i - P_2 L_j C_i - C_i^T L_j^T P_2 - k_j^T \Delta B_i^T P_2 - P_2 \Delta B_i(t) k_j$

$Y_{23} = P_2 \Delta A_i(t)$

$Y_{33} = A_r^T P_3 + P_3 A_r$

In order to solve the matrix inequality above we rewrite the matrix (18) in format of to matrices the constant term and the uncertainties term as follows:

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & -P_1 & P_1 B_{2i} \\ * & A_{22} & 0 & 0 & P_2 B_{2i} \\ * & * & A_{33} & P_3 & 0 \\ * & * & * & -\eta^2 I & 0 \\ * & * & * & * & -\eta^2 I \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & B_{13} & 0 & 0 \\ * & B_{22} & B_{23} & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix} \prec 0 \quad \text{With:} \tag{19}$$

$A_{11} = A_i^T P_1 + P_1 A_i + k_j^T B_i^T P_1 + P_1 B_i k_j + Q$

$A_{12} = -P_1 B_i k_j$

$A_{13} = P_1 A_i - P_1 A_r$

$A_{22} = A_i^T P_2 + P_2 A_i - P_2 L_j C_i - C_i^T L_j^T P_2$

$A_{33} = A_r^T P_3 + P_3 A_r$

$B_{11} = \Delta A_i^T(t) P_1 + P_1 \Delta A_i(t) + k_j^T \Delta B_i^T(t) P_1 + P_1 \Delta B_i(t) k_j$

$B_{12} = \Delta A_i^T(t) P_2 + k_j^T \Delta B_i^T(t) P_2 - P_1 \Delta B_i(t) k_j$

$B_{13} = P_1 \Delta A_i(t)$

$B_{22} = -k_j^T \Delta B_i^T P_2 - P_2 \Delta B_i(t) k_j$

$B_{23} = P_2 \Delta A_i(t)$

By introducing new variables $N = P^{-1}$, $X_j = k_j P_1^{-1}$ and $W_j = P_2 L_j$ after per-post-multiply the inequality (19) by $diag\,[P_1^{-1} \quad I \quad I \quad I \quad I]$ we obtain the following form:

$$\begin{bmatrix} \overline{A}_{11} & \overline{A}_{12} & \overline{A}_{13} & -N_1 & B_{2i}N \\ * & \overline{A}_{22} & 0 & 0 & P_2 B_{2i} \\ * & * & \overline{A}_{33} & P_3 & 0 \\ * & * & * & -\eta^2 I & 0 \\ * & * & * & * & -\eta^2 I \end{bmatrix} + \begin{bmatrix} \overline{B}_{11} & \overline{B}_{12} & \overline{B}_{13} & 0 & 0 \\ * & \overline{B}_{22} & \overline{B}_{23} & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix} \prec 0 \tag{20}$$

$$\begin{aligned}
\overline{A}_{11} &= N^T A_i^T + A_i N + X_j^T B_i^T + B_i X_j + NQN \\
\overline{A}_{12} &= -B_i X_j \\
\overline{A}_{13} &= A_i N - A_r N \\
\overline{A}_{22} &= A_i^T P_2 + P_2 A_i - W_j C_i - C_i^T W_j^T \\
\overline{A}_{33} &= A_r^T P_3 + P_3 A_r \\
\overline{B}_{11} &= N^T \Delta A_i^T(t) + \Delta A_i(t) N + X_j^T \Delta B_i^T(t) + \Delta B_i(t) X_j \\
\overline{B}_{12} &= N^T \Delta A_i^T(t) + X_j^T \Delta B_i^T(t) - \Delta B_i(t) X_j \\
\overline{B}_{13} &= \Delta A_i(t) N \\
\overline{B}_{22} &= -X_j^T \Delta B_i^T - \Delta B_i(t) X_j \\
\overline{B}_{23} &= \Delta A_i(t) N
\end{aligned}$$

Then using the uncertainties structure defined in (1) the term with uncertainties in (20) becomes:

$$\begin{aligned}
&\begin{bmatrix} H_i \\ H_i \\ 0 \\ 0 \\ 0 \end{bmatrix} F_1 \begin{bmatrix} E_{ai}N & 0 & E_{ai}N & 0 & 0 \end{bmatrix} + \begin{bmatrix} N^T E_{ai}^T \\ 0 \\ N^T E_{ai}^T \\ 0 \\ 0 \end{bmatrix} F_1^T \begin{bmatrix} H_i^T & H_i^T & 0 & 0 & 0 \end{bmatrix} \\
&+ \begin{bmatrix} H_i \\ H_i \\ 0 \\ 0 \\ 0 \end{bmatrix} F_2 \begin{bmatrix} E_{bi}X_j & -E_{bi}X_j & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} X_j^T E_{bi}^T \\ -X_j^T E_{bi}^T \\ 0 \\ 0 \\ 0 \end{bmatrix} F_2^T \begin{bmatrix} H_i^T & H_i^T & 0 & 0 & 0 \end{bmatrix}
\end{aligned} \tag{21}$$

The following lemmas are needed to propose LMI conditions:

Lemma 1.[] Given constant matrices H_i and E_{ai}, E_{bi} of appropriate dimensions , $\mu_i \succ 0$ and $F(t)$ satisfied $F^T(t)F(t) \le I$ the following inequality holds :

$$H_i F(t) E_{ai} + E_{ai}^T F^T(t) H_i^T \le \mu_i^{-1} H_i H_i^T + \mu_i E_{ai}^T E_{ai}$$

Lemma 2. [] shur complement the LMI:

$$\begin{bmatrix} Q(y) + \rho.I & S(y) \\ S^T(y) & R(y) \end{bmatrix} \prec 0$$

is equivalent to $R(y) \prec 0, Q(y) - S(y)R(y)^{-1}S^T(y) \prec -\rho.I$ where $Q(y) = Q^T(y), R(y) = R^T(y)$ and $S(y)$ depend affinely on y and ρ is a real number.

Remark 1. The parameter uncertainties $\Delta A_i(t), \Delta B_i(t)$ will represent the impossibility for exact mathematical model of a dynamic system due to the system complexity. The uncertainty has been widely used in many practical systems [11] [12].

By applying lemma .1 in term with uncertainties (21) we have:

$$\mu_1^{-1}\begin{bmatrix}H_i\\0\\0\\0\\0\end{bmatrix}\begin{bmatrix}H_i^T & 0 & 0 & 0 & 0\end{bmatrix}+\mu_1\begin{bmatrix}N^T E_{ai}^T\\0\\0\\0\\0\end{bmatrix}\begin{bmatrix}E_{ai}N & 0 & 0 & 0 & 0\end{bmatrix}$$
$$+\mu_2^{-1}\begin{bmatrix}0\\H_i\\0\\0\\0\end{bmatrix}\begin{bmatrix}0 & H_i^T & 0 & 0 & 0\end{bmatrix}+\mu_2\begin{bmatrix}0\\0\\N^T E_{ai}^T\\0\\0\end{bmatrix}\begin{bmatrix}0 & 0 & E_{ai}N & 0 & 0\end{bmatrix}$$
$$+\mu_3^{-1}\begin{bmatrix}H_i\\0\\0\\0\\0\end{bmatrix}\begin{bmatrix}H_i^T & 0 & 0 & 0 & 0\end{bmatrix}+\mu_3\begin{bmatrix}X_j^T E_{bi}^T\\0\\0\\0\\0\end{bmatrix}\begin{bmatrix}E_{bi}X_j & 0 & 0 & 0 & 0\end{bmatrix} \quad (22)$$
$$+\mu_4^{-1}\begin{bmatrix}0\\H_i\\0\\0\\0\end{bmatrix}\begin{bmatrix}0 & H_i^T & 0 & 0 & 0\end{bmatrix}+\mu_4\begin{bmatrix}0\\-X_j^T E_{bi}^T\\0\\0\\0\end{bmatrix}\begin{bmatrix}0 & -E_{bi}X_j & 0 & 0 & 0\end{bmatrix}$$

Then, replacing (22) and (20) in (19) and applying lemma.2, we get the LMI (11), this condition can be efficiently solved numerically by using a linear matrix inequality (LMI) framework, wich ends the proof.

4 Example and Simulation

To illustrate the proposed fuzzy robust tracking control condition, a control problem of a wind generator is considered.

By defining the state vector $x(t)=\begin{bmatrix}\theta_s & \Omega_r & \Omega_g & \beta\end{bmatrix}^T$ and the control signal as $u=[\beta_d,\Omega_z]^T$, The reference state vector to track is defined as $x_r(t)=\begin{bmatrix}\theta_{rs} & \Omega_{rr} & \Omega_{gr} & \beta_r\end{bmatrix}^T$ the fuzzy model of the WECS can be described as:

$$\begin{cases}\dot{x}(t)=A(z)x(t)+Bu(t)+B_2V(t)\\ y(t)=Cx(t)\end{cases} \quad (23)$$

Where:

$$A(z)=\begin{bmatrix}0 & 1 & -1 & 0\\ -\frac{K_s}{J_r} & -\frac{B_s}{J_r} & \frac{B_s}{J_r} & \frac{T_{r\beta}(z_0)}{J_r}\\ \frac{K_s}{J_g} & \frac{B_s}{J_g} & -\frac{(B_s+B_g)}{J_g} & 0\\ 0 & 0 & 0 & -\frac{1}{\tau}\end{bmatrix}$$

$$B=\begin{bmatrix}0 & 0\\ 0 & 0\\ 0 & \frac{B_g}{J_g}\\ \frac{1}{\tau} & 0\end{bmatrix} \qquad B_2=\begin{bmatrix}0\\ \frac{T_{rv}(z_0)}{J_r}\\ 0\\ 0\end{bmatrix}$$

$$C=\begin{bmatrix}0 & 0 & 1 & 0\end{bmatrix}$$

Where θ_s denotes the torsion angle, Ω_r is angular velocity of rotor, Ω_g is angular velocity of generator, K_s is the stiffness of the transmission, B_s is the damping of the transmission, J_r and J_g are the inertia of the rotor and generator, respectively.

T_r is the aerodynamic torque. β and β_d are the actual and desired pitch angles, respectively.

4.1 T-S Fuzzy Modeling:

The wind generator system is then described by the following four IF-THEN rules:

If β is F_1^1 and V is F_2^1 then $\begin{cases} \dot{x} = A_1 x + B_1 u + B_{21}\varphi \\ y = Cx \end{cases}$

If β is F_1^1 and V is F_2^2 then $\begin{cases} \dot{x} = A_2 x + B_2 u + B_{22}\varphi \\ y = Cx \end{cases}$

If β is F_1^2 and V is F_2^1 then $\begin{cases} \dot{x} = A_3 x + B_3 u + B_{23}\varphi \\ y = Cx \end{cases}$

If β is F_1^2 and V is F_2^2 then $\begin{cases} \dot{x} = A_4 x + B_4 u + B_{24}\varphi \\ y = Cx \end{cases}$

In order to obtain the best performance from this non linear system (23) the following T-S fuzzy model is given:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^{4} h_i(z(t))\left[(A_i + \Delta A_i(t))x(t) + (B_i + \Delta B_i(t))u(t) + B_{2i}\varphi(t)\right] \\ y(t) = \sum_{i=1}^{4} h_i(z(t))C_i x(t) \end{cases} \quad (24)$$

The parameter uncertainties $\Delta A_i(t), \Delta B_i(t)$ represent the impossibility for exact mathematical model of a dynamic system due to the system complexity. The wind generator is a complex system then the presence of uncertainties is possible.

Where:

$$\varphi(t) = V(t)$$

$$h_1(z) = F_1^1(\beta)F_2^1(V), \quad h_1(z) = F_1^1(\beta)F_2^1(V),$$
$$h_2(z) = F_1^1(\beta)F_2^2(V) \quad h_2(z) = F_1^1(\beta)F_2^2(V)$$
$$h_3(z) = F_1^2(\beta)F_2^1(V), \quad h_3(z) = F_1^2(\beta)F_2^1(V),$$
$$h_4(z) = F_1^2(\beta)F_2^2(V) \quad h_4(z) = F_1^2(\beta)F_2^2(V)$$

$$A_1 = \begin{bmatrix} 0 & 1 & -1 & 0 \\ -\frac{K_s}{J_r} & -\frac{B_s}{J_r} & \frac{B_s}{J_r} & \frac{T_{r\beta 1}}{J_r} \\ \frac{K_s}{J_g} & \frac{B_s}{J_g} & -\frac{(B_s + B_g)}{J_g} & 0 \\ 0 & 0 & 0 & -\frac{1}{\tau} \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 0 & 1 & -1 & 0 \\ -\frac{K_s}{J_r} & -\frac{B_s}{J_r} & \frac{B_s}{J_r} & \frac{T_{r\beta 2}}{J_r} \\ \frac{K_s}{J_g} & \frac{B_s}{J_g} & -\frac{(B_s + B_g)}{J_g} & 0 \\ 0 & 0 & 0 & -\frac{1}{\tau} \end{bmatrix}$$

$$A_2 = A_1, A_4 = A_3$$

$$B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \frac{B_g}{J_g} \\ \frac{1}{\tau} & 0 \end{bmatrix} B_{21} = B_{23} = \begin{bmatrix} 0 \\ \frac{T_{rv1}}{J_r} \\ 0 \\ 0 \end{bmatrix} B_{22} = B_{24} = \begin{bmatrix} 0 \\ \frac{T_{rv2}}{J_r} \\ 0 \\ 0 \end{bmatrix}$$

$C = [0 \quad 0 \quad 1 \quad 0]$

$B = B_1 = B_2 = B_3 = B_4$

$C = C_1 = C_2 = C_3 = C_4$

Numerical value:

$K_s = 1.566 * 10^6 N/m, \tau = 100\ ms$

$B_s = 3029.5 Nms/rad, B_g = 15.993 Nms/rad$

$J_r = 830000\ Kg.m^2, J_g = 5.9 Kg.m^2$

$T_{r\beta 1} = 723980, T_{r\beta 2} = 37607($

$T_{rv1} = 106440, T_{rv2} = 85370$

The control objective to attain is the best tracking of rated power while regulating rotor speed.

We consider output tracking control network-based T-S fuzzy system described by (24), we show the effectiveness of the proposed method by performing output tracking control for the T-S fuzzy system.

Solving LMI (11) by using the Toolbox LMI optimization algorithm, feedback and observer gain matrices can be obtained as:

$$k_1 = \begin{bmatrix} -0.0000 & -0.0000 & 0.0000 & 0.0000 \\ -0.0003 & -0.0000 & -0.0457 & -0.0000 \end{bmatrix}$$

$$k_2 = \begin{bmatrix} -0.0000 & -0.0000 & 0.0000 & 0.0000 \\ -0.0003 & -0.0000 & -0.0457 & -0.0000 \end{bmatrix}$$

$$k_3 = \begin{bmatrix} -0.0000 & -0.0000 & 0.0000 & 0.0000 \\ -0.0003 & -0.0000 & -0.0457 & -0.0000 \end{bmatrix}$$

$$k_4 = \begin{bmatrix} -0.0000 & 0.0000 & 0.0000 & -0.0000 \\ -0.0003 & -0.0000 & -0.0568 & -0.0000 \end{bmatrix}$$

$$L_1 = \begin{bmatrix} 0.0003 \\ 0.0000 \\ 0.0764 \\ 0.0000 \end{bmatrix}, L_2 = \begin{bmatrix} 0.0003 \\ 0.0000 \\ 0.0764 \\ 0.0000 \end{bmatrix}$$

$$L_3 = \begin{bmatrix} 0.0003 \\ 0.0000 \\ 0.0764 \\ 0.0000 \end{bmatrix}, L_4 = \begin{bmatrix} 0.0004 \\ 0.0000 \\ 0.0877 \\ 0.0000 \end{bmatrix}$$

The simulations results of LMI (11) show that the attenuation level is $\eta = 0.3442$ so, we can deduce that the designed T-S fuzzy controller ensures a good tracking performance and can to guarantee the stability.

For simulation, we consider the wind speed input profile

(17m/s < V < 31m/s) as given in Figure.1.

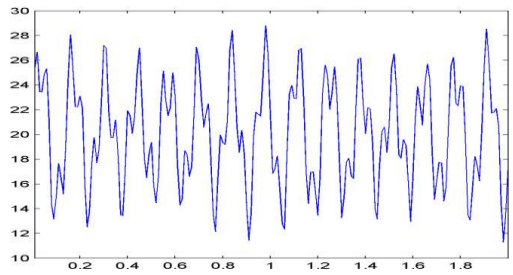


Figure. 1. The profile of wind input speed of 17w32 m/s

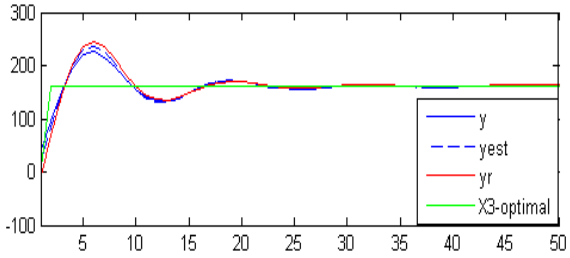


Figure. 2. Rotor speed (rd/s) with H∞ observer based

The above figure is presenting the speed rotor tracking. It is clear that the curve follws its reference for the different wind speed variations. Thus, we can deduce that the designed T-S fuzzy controller ensures a good tracking performance. The following figure shows the membership function .

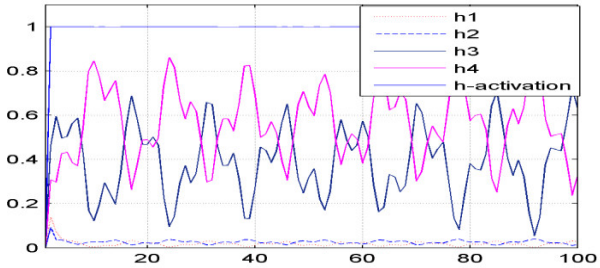


Figure. 3.membership function

5 Conclusion

In the present paper, we have proposed a sufficient condition for robust stabilization of uncertain non linear system with external disturbances .a robust fuzzy tracking controller and robust fuzzy observer are devllepoed and analyzed. The main result of this work is the quasi-LMI formulation that can be applied for control design of uncertain and disturbed wind turbine . Based on the T-S fuzzy model, a fuzzy observer based fuzzy controller is developed to reduce the tracking error as much as possible by minimizing the effect of the disturbance level.

REFERENCES

[1] (Takagi & Sugeno, 1985) Takagi, T., and Sugeno, M., Fuzzy identification of systems and its application to modelling and control". IEEE Trans. Syst., Man and Cyber, Vol.1115, pp. 116-132, 1985

[2] ShaochengTong,TaoWang,Han-Xiong Li "Fuzzy robust tracking control for uncertain nonlinear systems " (2002)

[3] K. Rai Lee, E. Tae Jeung, H. Bae Park"Robust fuzzy H∞control for uncertain nonlinear systems via state feedback: an LMI approach"(2001)

[4] B.Mansouri, N. Manamanni, A.Hamzaoui and J. Zaytoon "Tracking control for uncertain Takagi Sugeno fuzzy systems with external disturbances "(2005)

[5] S. Bououden, M. Chadli,, S. Filali, A. El Hajjaji "Fuzzy model based multivariable predictive control of a variable speed wind turbine: LMI approach"(2012)

[6] K. Tanaka, and K. O. Wang. Fuzzy Control Systems Design and Analysis: A linear Matrix Inequality Approach.John Wiley & Sons, Inc., 2001

[7] Boyd, S., El Ghaoui, L., Feron, E., &Balakrishnan, V. (1994). Linearmatrix inequalities in system and control theory. In SIAM Studies in Applied Mathematics: vol. 15. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM)

[8] B. Mansouri, N. Manamanni, K. Guelton, A. Kruszewski, T.M. Guerra "Output feedback LMI tracking control conditions withH1criterion for uncertain and disturbed T–S models"(2008)

[9] N. Harrabi, M. Kharrat, M. Souissi, A. Aitouche "Maximum Power Point Tracking of a Wind Generation System Based on T-S Fuzzy Model"(2015)

[10] C.Tseng, B.Chen, and H.Uang "Fuzzy Tracking Control Design for Nonlinear Dynamic Systems via T–S Fuzzy Model"(2001)

[11] D.Zhang, Q. Han and X. Jia "Tracking Control for Network-Based T-S Fuzzy Systems With Asynchronous Constraints"(2012)

[12] S. Bezzaoucha, B. Marx, D. Maquin, J. Ragot "Model Reference Tracking Control for Nonlinear Systems described by Takagi-Sugeno Structure"(2013)

[13] K. Ben Meziane, I. Boumhidi "An Interval Type-2 Fuzzy Logic PSS with the optimal H ∞ tracking control for multi-machine power system"(2015)

[14] A. Abdelkrim, C. Ghorbel, M. Benrejeb "LMI-based tracking control for Takagi-Sugeno fuzzy model"(2010)

[15] F.Khaber, A.Hamzaoui, K. Zehar "State Feedback Controller Design via Takagi-Sugeno fuzzy Model : A Linear Matrix Inequalities Approach"

[16] J. Zhang, M.Fei, T. Yang and Y. Tan "Robust Fuzzy Tracking Control of Nonlinear Systems with Uncertainty Via T-S Fuzzy Model"(2006)

[17] Allouche,,M. Chaabane,M. Souissi,D. Mehdi ,F. Tadeo "STATE FEEDBACK TRACKING CONTROL FOR INDIRECT FIELD-ORIENTED INDUCTION MOTOR USING FUZZY APPROACH"(2013

[18] Chang-Hua Lien"An efficient method to design robust observer-based control of uncertain linear systems "(2004)

[19] M. Chadli and A. El Hajjaji "Wind energy conversion systems control using T-S fuzzy modeling"(2010

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Integration of the Cloud Environment in E-learning Systems

Karim Dahdouh, Lahcen Oughdir, Ahmed Dakak
LSI, Department of Mathematics, Physics and Informatics
University of Sidi Mohamed Ben Abdellah, FPT,Taza, Morocco
Karim.dahdouh@usmba.ac.ma, ahcen.oughdir@usmba.ac.ma ,ahmed.dakak@ usmba.ac.ma

ABSTRACT

Nowadays, e-learning systems have known a major revolution, especially with the emergence of new information and communication technologies and the enormous growth in the number of learners, educational content and resources. which causes many challenges in terms of optimizing resources such as compute, storage and communication, highlighting the need to use a model that meets evolving requirements and cost control. In this context, cloud computing is proving to be a promising model because of its permanent scalability and the efficient use of resources. This paper presents the characteristics of current e-learning systems, then analyzes the concept of cloud computing and describes the architecture of it by combining the features of e-learning. We propose an approach of using and harnessing Cloud computing services in e-learning systems to build an e-learning environment based on cloud computing.

Keywords : E-learning, Cloud Computing, Cloud services, e-learning cloud, e- learning based cloud, Learner, Iaas,Saas, Paas.

1 Introduction

At present, most of the conventional education forms are becoming not being suitable for requirements of social progress and educational development and not being able to catch up with the changes of learning demand Which continues to increase especially with the emergence of new information technologies and the enormous growth in the number of learners, The educational content, the services that can be offered and the resources available, the dimensions of the e-learning system grow at an exponential rate. Challenges With regard to optimization of resource requirements, storage and communication requirements and the processing of requests for simultaneous competition, it is necessary to use a platform that meets the evolving requirements and the control costs.

However, in traditional web-based e-learning mode, system construction and maintenance are located in interior of educational institutions or enterprises, which results in a lot of problems existed, such as a lot of investment needed, but without capital gains to return, without development potential and staying power. Cloud computing is becoming an attractive technology due to its dynamic scalability and effective usage of the resources; it can be utilized under circumstances where the availability of resources is limited.

DOI: 10.14738/tmlai.54.2973
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2973

The new information technologies have changed the way of teaching and learning. Being the state of the art in education means that the institutions providing e-learning should offer new opportunities for self-regulated e-learning consistent with the new expectations and needs of nowadays digital learners.

Combining the features of traditional learning systems with the cloud-based ones will give the organizations a powerful tool that is cost effective, easy to maintain, secure and engaging learners and teachers. But until now the research applying cloud computing to e-learning is not significantly reported. In order to give a full play for the advantages of cloud computing, in this paper, we tried to attach cloud computing to e-learning, build an e-learning cloud, and made an active research and exploration for it.

2 Cloud Computing

2.1 General context

The remarkable development of cloud computing in recent years is increasingly sparking the interest of Internet and IT users seeking to derive the greatest benefit from the services and applications available on line via the web in service-on-demand mode with per-usage billing.

Cloud Computing is a new paradigm to organize and manage ICT resources. There are various definitions of cloud computing, one of which is the definition according to The National Institute of Standards and Technology (NIST) which defines cloud computing as “model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” [26].

In this context Praveena and Betsy [17] have described the application of Cloud in universities. Delic and Riley [18] assessed the current state of enterprise knowledge management and how it would turn into a more global, dependable and efficient infrastructure with Cloud computing. They have discussed architecture as well as applications. Cloud computing attributes can be visualized from the following comparison:

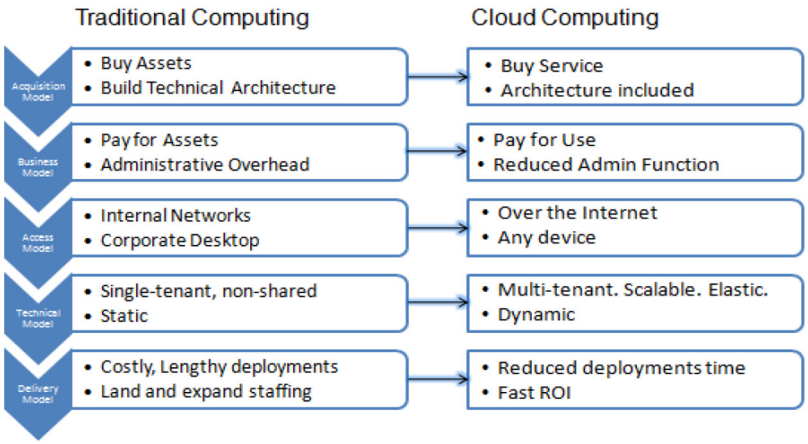


Figure. 1 .Comparison between traditional computer and cloud

2.2 Cloud computing deployment models

The principle of the cloud is to pool the resources of infrastructures and services in order to optimize their use and to realize economies of scale. According to the National Institute of Standards and Technology (NIST) [26], four of the most common cloud deployment models are:

- Public cloud: The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.
- Private cloud: The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.
- Community cloud: The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.
- Hybrid cloud: The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

A computing environment can use a multitude of clouds, whether public, private, community or hybrid, as shown in Fig. 2.

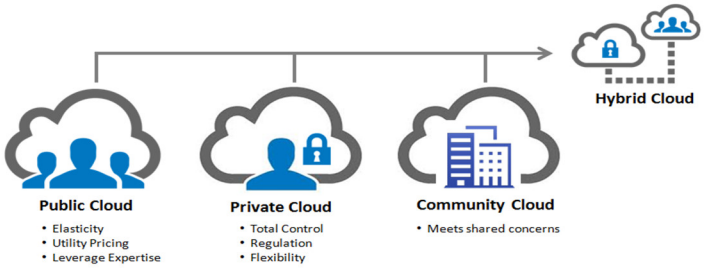


Figure 2 .Cloud Computing deployment models

2.3 Cloud computing services models

Cloud service models describe how cloud services are made available to clients. Most fundamental service models include a combination of IaaS (infrastructure as a service), PaaS (platform as a service), and SaaS (software as a service). These service models may have synergies between each other and be interdependent – for example, PaaS is dependent on IaaS because application platforms require physical infrastructure Fig. 3.

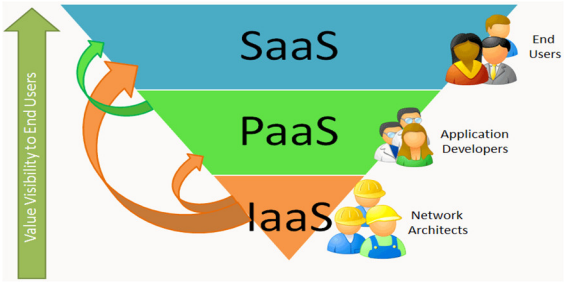


Figure 3 .Cloud computing services models

Infrastructure as a Service (IaaS): The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

Software as a Service (SaaS): The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

Platform as a Service (PaaS): The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

3 From traditional e-learning network to cloud e-learning

E-learning is an Internet-based learning process, using Internet technology to design, implement, select, manage, support and extend learning, which will not replace traditional education methods, but will greatly improve the efficiency of education. As e-learning has a lot of advantages like flexibility, diversity, measurement, opening and so on, it will become a primary way for learning in the new century as in Fig. 4 [27].

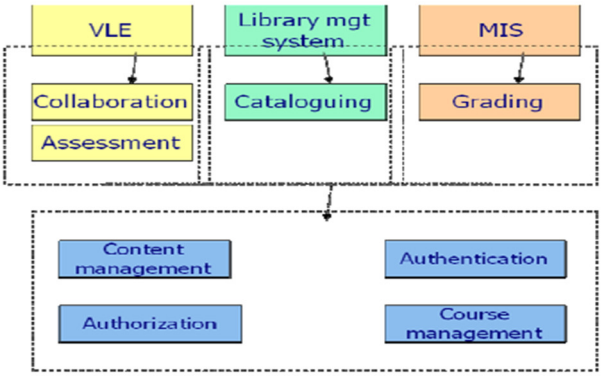


Figure 4. Architecture of a simplified Learning System

Mendez [19] illustrates that in traditional web-based learning mode, system construction and maintenance are located inside the educational institutions or enterprises, which led to a lot of problems, such as significant investment needed but without capital gains for them, which leads to a lack of development potential. In contrast, cloud based e-learning model introduces scale efficiency mechanism, i.e. construction of e-learning system is entrusted to cloud computing suppliers, which can make providers and users to achieve a win-win situation. The cloud-based environment supports the creation of new generation of e-learning systems, able to run on a wide range of hardware devices, while storing data inside the cloud.

Ouf [20] has presented an innovative e-learning ecosystem based on cloud computing and Web 2.0 technologies. The article analyses the most important cloud-based services provided by public cloud computing environments such as Google App Engine, Amazon Elastic Compute Cloud (EC2) or Windows Azure, and highlights the advantages of deploying E-Learning 2.0 applications for such an infrastructure. The authors also identified the benefits of cloud-based E-Learning 2.0 applications (scalability, feasibility, or availability) and underlined the enhancements regarding the cost and risk management.

Chandran [21] focused on current e-learning architecture model and on issues in current e-learning applications. The article presents the Hybrid Instructional Model as the blend of the traditional classroom and online education and its customization for e-learning applications running on the cloud computing infrastructure. The authors underline the e-learning issues, especially the openness, scalability, and development/customization costs. The existing e-learning systems are not dynamically scalable and hard to extend – integration with other e-learning systems is very expensive. The article proposed the hybrid cloud delivery model that can help in fixing the mentioned problems.

In this article a new paradigm is highlighted in educational area by introducing the cloud computing in order to increase the scalability, flexibility and availability of e-learning systems. The authors have evaluated the traditional e-learning networking model, with its advances and issues, and the possibility to move the e-learning system out of schools or enterprises, inside a cloud computing infrastructure. The separation of entity roles and cost effectiveness can be considered important advantages. The institutions will be responsible for the education process, content management and delivery, and the vendor takes care of system construction, maintenance, development and management. The e-learning system can be scaled, both horizontally and vertically, and the educational organization is charged according to the number of used servers that depends on the number of students as in Fig. 5 [24].

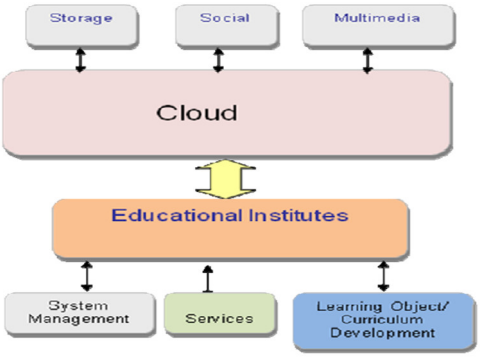


Figure 5 . Modified E-learning System Architecture.

The e-learning cannot completely replace teachers; it is only an updating for technology, concepts and tools, giving new content, concepts and methods for education, so the roles of teachers cannot be replaced. The teachers will still play leading roles and participate in developing and making use of e-learning cloud. The blended learning strategy should improve the educational act. Moreover, the interactive content and virtual collaboration guarantee a high retention factor.

On the other hand, E-learning based cloud is a migration of cloud computing technology in the field of e-learning, which is a future e-learning infrastructure, including all the necessary hardware and software computing resources engaging in e-learning. After these computing resources are virtualized, they can be afforded in the form of services for educational institutions, students and businesses to rent computing resources. E-learning based cloud approach is shown in Fig.6.

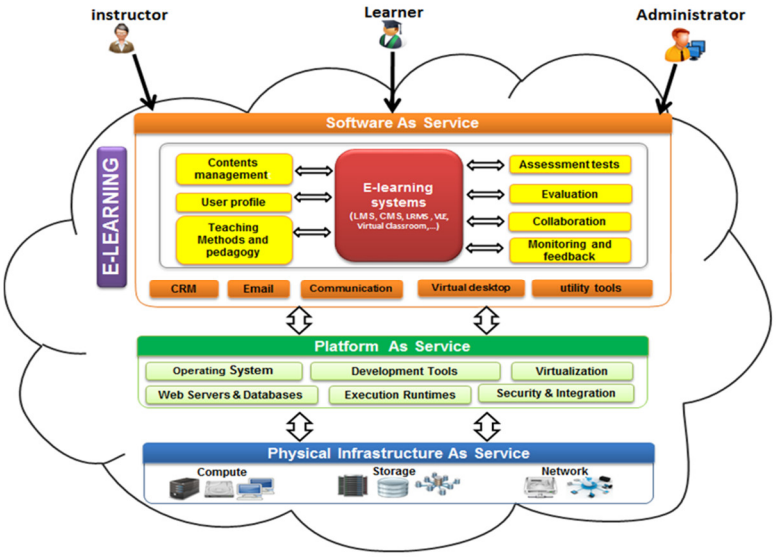


Figure 6 . E-learning based Cloud Approach.

The proposed e- learning cloud architecture can be divided into the following layers: Physical infrastructure layer as a dynamic and scalable physical host pool, platform as a service layer provides resources to build applications, Software as a service layer consist of complete applications that do not require development it that provides with content production, content delivery, virtual laboratory, collaborative learning, assessment and management features.

3.1 Infrastructure layer

Infrastructure layer is composed of information infrastructure and teaching resources. Information infrastructure contains Internet/Intranet, system software, information management system and some common software and hardware; teaching resources is accumulated mainly in traditional teaching model and distributed in different departments and domain. This layer is located in the lowest level of cloud service middleware, the basic computing power like physical memory, CPU, memory is provided by the layer. Through the use of virtualization technology, physical server, storage and network form virtualization group for being called by upper software platform. The physical host pool is dynamic and scalable, new physical host can be added in order to enhance physical computing power for cloud middleware services. The following Fig. 7 depicts this in a clearer view [27].

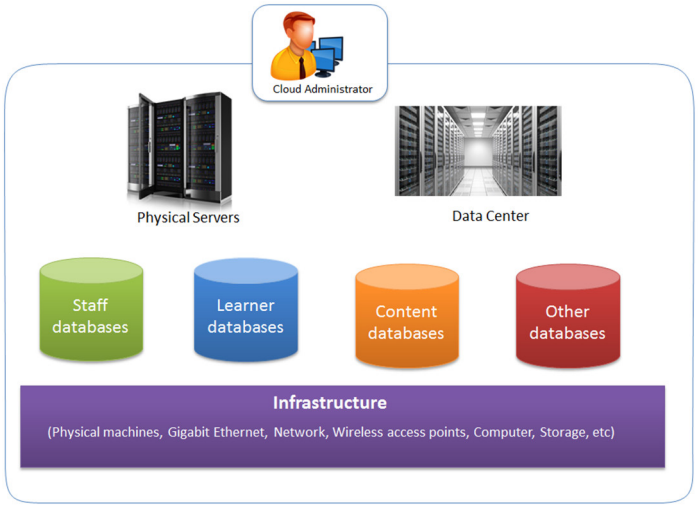


Figure. 7 . Proposed Infrastructure layer

3.2 Platform as a service layer

This layer is generally composed of operating system (Windows, UNIX, etc.), the development Tools, the execution runtimes, the virtualization, the web servers, the databases and the security.

The provider supplies a platform of software environments and application programming interfaces (APIs) that can be utilized in developing cloud applications. Naturally, the users of this class of systems are developers who use specific APIs or programing language to build, test, deploy, and tune their applications on the cloud platform.

At this layer customers do not manage their virtual machines, there is no need to manage an operating system, and it is the supplier who occupies these tasks. E-learning professionals merely create their own programs which are hosted by the platform services they are paying for.

3.3 Software as a service layer

The third and final layer of the cloud is Software as a Service, or Application as a service. This layer is the one that learner and tutor most likely to interact with in the learning process, and it is almost always accessible through a web browser. Any application hosted on a remote server that can be accessed over the Internet is considered a SaaS.

Software as a service layer is the specific applications of integration the teaching resources in the cloud computing model, including interactive courses and sharing the teaching resources. The interactive programs are mainly for the teachers, according to the learners and teaching needs, taken full advantage of the underlying information resources after finishing made, and the course content as well as the progress may at any time adjust according to the feedback, and can be more effectiveness than traditional teaching. Sharing of teaching resources include teaching material resources, teaching information resources (such as digital libraries, information centers), as well as the full sharing of human resources. This layer mainly consists of content production, educational objectives, Teaching Methods and pedagogy, content delivery technology, contents management, assessment and Monitoring and feedback.

In the classic e- learning model, teachers assign teaching tasks, conduct regular lectures, or train students skills. The students attend the online autonomous learning act and cooperative learning sessions, or accomplish teachers' assignments. But in the proposed architecture teachers also answer students' questions and offer essential teaching to major and difficult points. In addition, teachers can also use multimedia to enhance teaching content. Students work out their own learning plans, determining learning methods autonomously. They conduct on-line autonomous learning when they study each unit, finish its test via Internet and do some statistics to the test results. Teachers also encourage students to cooperate with each other to finish simple learning tasks or complex group-based projects. Through cooperative learning, students cannot only acquire knowledge, their team spirit and coordination will also be fostered, skills in dealing with people will be improved and abilities to express themselves will be enhanced. Thus the learning and teaching will be more interactive which is the demand of the age. The interactive mode of the proposed architecture is furnished in the Fig. 8 [27].

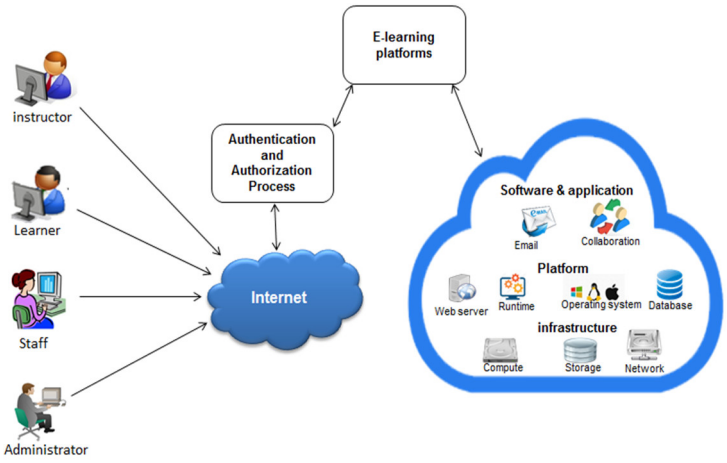


Figure 8 . Interactive mode of the proposed approach

4 Expected benefit from E-learning based cloud

There are several advantages of e-learning implementation of cloud computing technology, which include:

a) Powerful computing and storage capacity: Cloud based E-learning architecture locates the computing and data in a large number of distributed computers, the sea of clouds in the tens of thousands of computers to provide powerful computing power and huge data storage space, puts the "cloud" as a service available to students via the Internet.

b) It lowers the cost: users need not to have high end configured computers to run e-learning applications. Through cloud they can run the applications through their mobile phones, PC, tablet PC with internet connection.as the data created is accessed in the cloud, the user need not to spend more money on storage devices for large memory in local machines. And organizations also need to pay for per use, so it's cheaper.

c) High availability: Through the integration of mass storage and high-performance computing power, this system can provide a higher quality of service. Cloud computing system can automatically detect the node failure and exclude it, do not affect the normal operation of the system.

d) High security: In the cloud computing model, data is storied intensively. Relying on one or more data center, the managers manage the unified data, allocate the resources, balance load, deploy the software, control security, and do the reliable real time monitoring, thus guarantee the users' data security to the greatest possible degree.

e) Virtualization: Virtualization is the most important characteristics of this type of architecture. Each application deployment environment and physical platform is not related. It is managed, expensed, migrated, and backup through virtualization platform. It put the underlying hardware, including servers, storage and networking equipment, comprehensive virtualization, in order to build a resources pool of shared, distributed on-demand.

f) Instant software updates: The software are automatically updated in cloud computing. So the users can get the updates instantly.

g) Improved compatibility with document formats: Since some file formats and fonts do not open properly in some PCs/mobile phones, the cloud powered E-learning applications do not have to worry about those kinds of problems. As the cloud based E-learning applications open the file from the cloud.

h) Benefits for students: Students get more advantage of e leaning based on cloud as they can enhance their skills by taking online courses, attending online exams, getting feedback from their instructor, and can send their assignments and projects online to their teachers.

i) Benefits for teachers: Teachers also get numerous benefits of cloud based e-learning. Teachers can prepare for online tests for students, interaction and better resources for students through content management, assessment tests, homework and projects undertaken by students, feedback and communication through the forum online.

The major advantage of the proposed approach is that it aims at providing easy access to costly software running on high performance processors to rural students at institutions which lack considerable facilities. Considerable investment would be required to implement this architecture, but the benefits would easily justify the cost.

5 Conclusion and Future Work

This paper proposes an approach for integrate cloud computing services in e-learning systems. The implementation of e-learning is now generally constructed separately by each institution. The deployment of e-learning platforms is costly because it takes the cost for provision of infrastructure, systems development, and hiring IT staff to maintain and enhance e-learning systems. Cloud computing is the most promising computing model using from last decade in distributed computing. With the implementation of e-learning in a cloud environment, educational institutions no longer have to pay for the provision of

infrastructure because it has been provided by the cloud service provider and institutions that wish to use it only pay according to the usage. The aim of our work was to propose an approach which will be using the services of Cloud Computing in e-learning systems. Mainly, we have presented the characteristics of current e-learning systems, then we have analyzed the concept of cloud computing and describes the architecture of it by combining the features of e-learning, and finally we have discussed the expected benefits of the proposed approach. Future research will include a study regarding the implementation strategy for migration to the proposed approach based on cloud.

REFERENCES

[1] F. Jian, "Cloud computing based distance education outlook", China electronic education, 2009.10, Totally 273, pp.39-42.

[2] R. Hua, "Teaching Information System Based on Clou d Computing", Computer and Telecommunications, 2010.02, pp. 42-43.

[3] Y. Juan, S. Yi-xiang, "The Initial Idea of New Lear ning Society which Based on Cloud Computing", Modern Educational Techn ology, Vol.20, No.1, 2010, pp.14-17.

[4] T. Jian, F. Lijian, G. Tao, "Cloud computing-based Design of Network Teaching System", Journal of TaiYuan Urban Vocation al college, Mar. 2010, pp.159-160.

[5] Y. Zhongze, "The basic principles of cloud computin g and its impact on education", Satellite TV and Broadband Multimedia, 2010.6, pp.67-70.

[6] W. Xiaomei, J. Xiaoqiang, "Cloud computing on the Impact of Higher Education", Science & Technology Information, 2010. 10, pp.397-398.

[7] Z. Zhong-ping, L. Hui-cheng，"The Development and Exploring of E- Learning System on Campus Network"，Journal of Shanxi Teacher's University (Natural Science Edition), Vol .18, No.1, Mar. 2004, pp.36-40.

[8] W. Jianmin, "Campus Network's E-learning Mode", New Curriculum Research, 2007.08, pp.84-86.

[9] Y. Wei, Y. Rong, "Research of an E-learning System Model Based on Agent", Computer Engineering and Applications, Nov. 2004, pp.156-158.

[10] A. Gladun, J. Rogushina, F. Garcı´a-Sanchez, R. Martı´nez-Be´jar, J. Toma´s Ferna´ndez-Breis, "An application of intelligent techniqu es andsemantic web technologies in e-learning environments", Expert Systems with Applications 36, 2009, 922-1931.

[11] Y. Li, S. Yang, J. Jiang, M. Shi, "Build grid-enab led large-scale collaboration environment in e-learning grid", Expe rt Systems with Applications 31,2006, 742-754.

[12] Z. Chengyun, "Cloud Security: The security risks o f cloud computing, models and strategies", Programmer, May.2010, pp.71 -73.

[13] B. Hayes, "Cloud computing," Comm. Acm, vol. 51, no. 7, pp. 9– 11, 2008.

[14] E. Tuncay, "Effective use of Cloud computing in educational institutions," Procedia Social Behavioral Sciences, p. 938–942, 2010.

[15] R. Buyya, C.S. Yeo & S.Venugopal, "Market-oriented Cloud computing: Vision, hype, and reality of delivering IT services as computing utilities," 10th Ieee Int. Conf. High Performance Comput. Comm., p. 5–13, 2009.

[16] M. Lijun, W.K. Chan & T.H. Tse, "A tale of Clouds: Paradigm comparisons and some thoughts on research issues," Ieee Asia-pasific Services Comput. Conf., Apscca08, pp. 464–469, 2008 .

[17] K. Praveena& T. Betsy, "Application of Cloud Computing in Academia," Iup J. Syst. Management, vol. 7, no. 3, pp. 50–54, 2009.

[18] K.A. Delic & J.A. Riley, "Enterprise Knowledge Clouds," Next Generation Km Syst. Int. Conf. Inform., Process, Knowledge Management, Cancun, Mexico, pp. 49–53, 2009.

[19] J. A. Mendez and E. J. González, “Implementing Motivational Features in Reactive Blended Learning: Application to anIntroductory Control Engineering Course“, IEEE Tran sactions on Education, Volume: PP, Issue: 99, 2011.

[20] S. Ouf, M. Nasr, and Y. Helmy, „An Enhanced E-Lear ning Ecosystem Based on an Integration between Cloud Computing and Web2.0”, Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pages 48-55, 2011.

[21] D. Chandran and S. Kempegowda, „Hybrid E-learning Platform based on Cloud Architecture Model: A Proposal”, Proc. Int ernational Conference on Signal and Image Processing (ICSIP), pages 534-537, 2010.

[22] L. Huanying, “Value and understanding for cloud com puting based on middleware”, Programmer, 2010.05. pp.68,69.

[23] F. feng, “Cloud-based IT infrastructure of next-gen eration telecom”, Mobile Communications, 2010, No. 8, pp.76-79.

[24] Md. Anwar Hossain Masud, Xiaodi Huan, “An E-learning System Architecture based on Cloud Computing”, World Academy of Science, Engineering and Technology 62 2012.

[25] H. Xin-ping, Z. Zhi-mei， D. Jian, “Medical Informatization Based on Cloud Computing Concepts and Techniques”, Journal of Medical Informatics, 2010, Vol.31, No.3, pp.6-9.

[26] NIST, Mell, P. and T. Grance, “The NIST Definition of Cloud Computing”, 2011, p. 6.

[27] Hossain , M. M. A. , & Huang, X. (2012). "An e -learning system architecture based on cloud computing". International Scholarly and Scientific Research & Innovation,6(2),736-740.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Technical Data Extraction and Representation in Expert CAPP System

Rafik Elaakil, Rechia Ahmed, Abdelilah El Mesbahi, Oussama Jaider
Department of Mechanical Engineering, Faculty of Sciences and Technics of Tangier, Ancienne Route de l'Aéroport, Km 10, Ziaten. BP : 416. Tanger - Maroc
rafik.elaakil@gmail.com,rechia@hotmail.com

ABSTRACT

Computer-aided process planning (CAPP) is an essential interface for linking design and manufacturing processes, the purpose of CAPP is to transform a part design specification obtained from CAD system into a sequence of machining operations which can be used by CAM system involving the selection of necessary manufacturing processes in order to transform a designed part into a final part economically and competitively.This paper focus on the extraction of tolerances (GD&T) and their association to the related features embedded into an expert CAPP system.

Keywords: CAD/CAM, CAPP, machining operation sequencing, Expert system, knowledge bases, Artificial intelligence

1 Introduction

Computer Aided Process Planning is an important activity in an intelligent manufacturing system. Several techniques have been proposed, implemented and tested in the context of generative CAPP; it consists of several activities including machining feature recognition, selection of machining operation type, selection of manufacturing resource, setup planning and operation sequencing.

The objective of precedence-constrained sequencing problem is to locate the optimal sequence dealing with time and costs, so that the resulting operation sequence could satisfy those constraints established by geometric technological, economical requirements, and then generate the optimal or near optimal operation sequences.

The extraction of tolerances data is an essential key that help to identify the basic constraints of the related input part. To achieve a complete representation for a process-planning model, the tolerances data (GD&T) needs to be completely extracted and represented through the input data file.

The development of a robust methodology for extracting this data for usea knowledge-based expert system, incorporating an artificial intelligence algorithm is the main thrust of the work reported in this paper.

DOI: 10.14738/tmlai.54.2974
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2974

2 Related researches work and review

Form features are not sufficient so that the part will be manufactured at a reasonable cost and will perform its intended function. Most researchers have concentrated on geometric data extraction, [1] describes a method for the generation of tolerance specifications from product data, for each feature one or more geometric controls will have to be identified, [2]presents a way to interpret the Geometric Dimensioning and Tolerancing (GD&T) specifications in STEP for tolerance analysis by utilizing OntoSTEP plugin [3] which consists of translating the APs defined in the EXPRESS language [4],and the actual instance files represented using the ISO 10303-21 (or simply part21)format [5].

[6] presents an on-machine measurement data model based on STEP-NC, this model represents tasks in process plan in an object-oriented, containing information on geometry and tolerances.

3 ICAPP-TURN Architecture

A main objective in modern design and advanced machining process is to achieve a significant reduction in product lead time through seamless integration between the various design and process planning activities. The concept of using component features to integrate a design system and a manufacturing system has been a major research direction in recent year, [7] describes the characteristics of an existing expert systems building tools and discusses the use and the importance of these basic characteristics from a perspective of computer aided process planning (CAPP) requirements in order to automate the process planning functions.

The framework of the proposed system is decomposed into several modules, Figure 1 show the conceptual model that defines the structure of ICAPP-TURN.

The process start from the knowledge acquisition step that consists of structuring and organizing knowledge from one source, usually human experts, so it can be used and represented into the knowledge representation module.

The knowledge database represents a centralized repository to store constraints and rules provided by the knowledge representation module, in other side, the technical data information is retrieved through the integration module that extract the technical information of such part, which can be represented to the system across the data representation module that maps this information into technical objects.

The core system is represented as an inference engine which helps to determine the optimal sequences plan using an

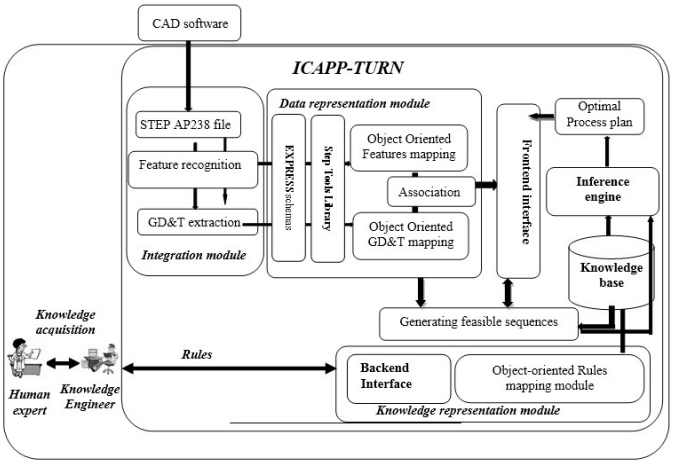


Figure1: Architecture of ICAPP-TURN

Artificial intelligence method, this module interacts with the Knowledge database.

4 Technical Data Integration and Representation modules

The integration module consists of recognizing manufacturing features and tolerances from a STEP AP238 file, then associate each tolerance to the related feature through the data representation module.

[8] has developed an automatic feature recognition system for rotational parts which adopted a STEP AP203 file as input to the system.

STEP AP203 is an application protocol of the Standard for the Exchange of Product model data (STEP), defined as the international standard ISO 10303 [9]. For the fact that the definitions of geometric and topological entities of the STEP AP203 are included in the AP238, we have used this feature recognition system to detect features from a STEP AP238, generated by STEP-NC Machine[ ]by applying this recognition, the system can recognize multiple features types. Figure 2 show an input part to our system. The feature recognition system recognizes in this case one shoulder, which is composed by 3 advanced face, one planar surface and two cylindrical surfaces:

#235=ADVANCED_FACE('Corps principal',(#171,#175),#179,.T.);

#47=ADVANCED_FACE('Corps principal',(#48),#82,.T.);

#87=ADVANCED_FACE('Corps principal',(#88),#82,.T.);

In the following section, we will explain how a type of tolerance such as surface roughness (named as SURFACE_TEXTURE_PARAMETER) is extracted from STEP AP238 file, and then associated to the corresponding feature.

As shown in Fig. 2, the face with the attribute #235 has a roughness of 1,6um as a value; this tolerance is embedded in the STEP AP238 data file of the example part asshown in Figure 3.

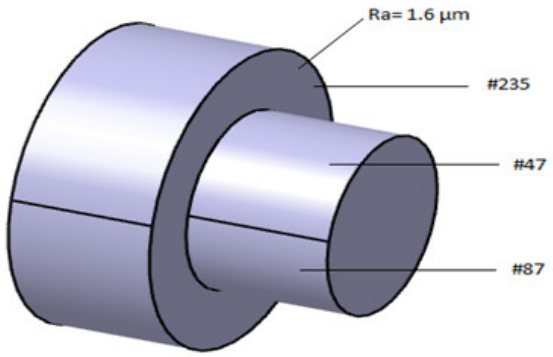


Figure2: An input example part that contains a toleranced feature

```
/*************************************************
 * Application object: CALLOUT (#231)
 * ITS_WORKPIECE: #231, #22, #21
 * EXPLICIT_REPRESENTATION [*]: #231, #232, #233, #234, #235
 * ITS_ID: #231, ['']
 */
#231=SHAPE_ASPECT('',#22,.F.);
#232=PROPERTY_DEFINITION('',#231);
#233=PROPERTY_DEFINITION_REPRESENTATION(#232,#234);
#234=REPRESENTATION('',(#235),#35);
#235=ADVANCED_FACE('Corps principal',(#171,#175),#179,.T.);
/*************************************************
 * Application object: SURFACE_TEXTURE_PARAMETER (#236)
 * ITS_VALUE: #236, #237
 * APPLIED_TO: #236, #238, #239, #231
 * MEASURING_METHOD: #236, #240, ['ISO 4287']
 * PARAMETER_NAME: #236, #237, ['Ra']
 */
#236=SURFACE_TEXTURE_REPRESENTATION('',(#237,#240),$);
#237=(
LENGTH_MEASURE_WITH_UNIT()
MEASURE_REPRESENTATION_ITEM()
MEASURE_WITH_UNIT(LENGTH_MEASURE(1.6),#241)
REPRESENTATION_ITEM('Ra')
);
#238=PROPERTY_DEFINITION_REPRESENTATION(#239,#236);
#239=PROPERTY_DEFINITION('',#231);
#240=DESCRIPTIVE_REPRESENTATION_ITEM('measuring method','ISO 4287');
```

Figure 3: A partial STEP AP238 text file of the definition of roughness of the part of Fig.2

In figure 3, the CALLOUT ( ) entity defines a callout identifying a particular face (ADVANCED_FACE) on the

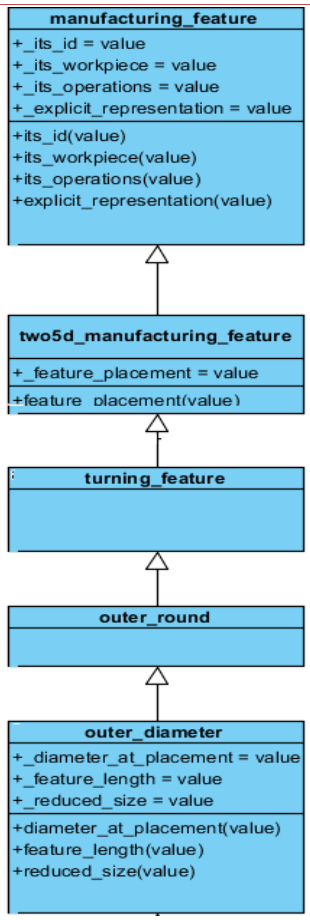


Figure 4: The UML diagram of the class outer_diameter

part that can have some properties (e.g. surface roughness). The SURFACE_TEXTURE_PARAMETERentity defines surface roughness to one or more faces; it has an attribute denoted as ITS_VALUE: #236, #237, where #237 is a pointer to the value and the type of surface roughness. The SURFACE_TEXTURE_PARAMETER ( ) entity has a second attribute denoted as APPLIED_TO: #236, #238, #239, #231, where #231 is a pointer to the callout of #231.

This means that the SURFACE_TEXTURE_PARAMETER is applied to the ADVANCED_FACE with the pointer #235.

Thus, following these indicators brings us allocate the surface roughness, its value and the advanced face to which thisroughness is applied using definedfunctions. The feature recognition module recognizes each feature and represents it as an object, for which The ADVANCED_FACE (that compose the feature) are known. This leads us to specify the feature for which the roughness has been applied to, however there is no association between the feature and the tolerance in term of objects. In fact, STEP Tools Library used in this module[11], generate an oriented object classes based on EXPRESS schema; the outer_diameter class doesn't have an attribute representing the tolerance applied to the concerned feature.

Figure 4 shows the class representing the outer_diameter feature according to ISO 14649, and the upstream classes from which the outer_diameter class inherits attributes.

As shown in Fig.4, the outer_diameter class doesn't have an attribute that represents the tolerance applied to the concerned feature. The solution proposed in this paper concerns the creation of an association between the parent class ofthe recognized features and the classes representing the tolerance, which means that an instance of the class outer_diameter for example is composed of zero or a list of instances (zero-to-Many Mapping) for each type of tolerance, whether it is a surface_texture_parameter or another tolerance type defined in STEP AP238.

Figure 5 shows the new association between turning_feature parent class and the new classes that can represent any type of tolerance. Taking the fact that tolerances classes have become part of the turning_feature class, each manufacturing feature (for example the outer_diameter class) for which the turning_feature class is parent will inherits tolerances as objects.

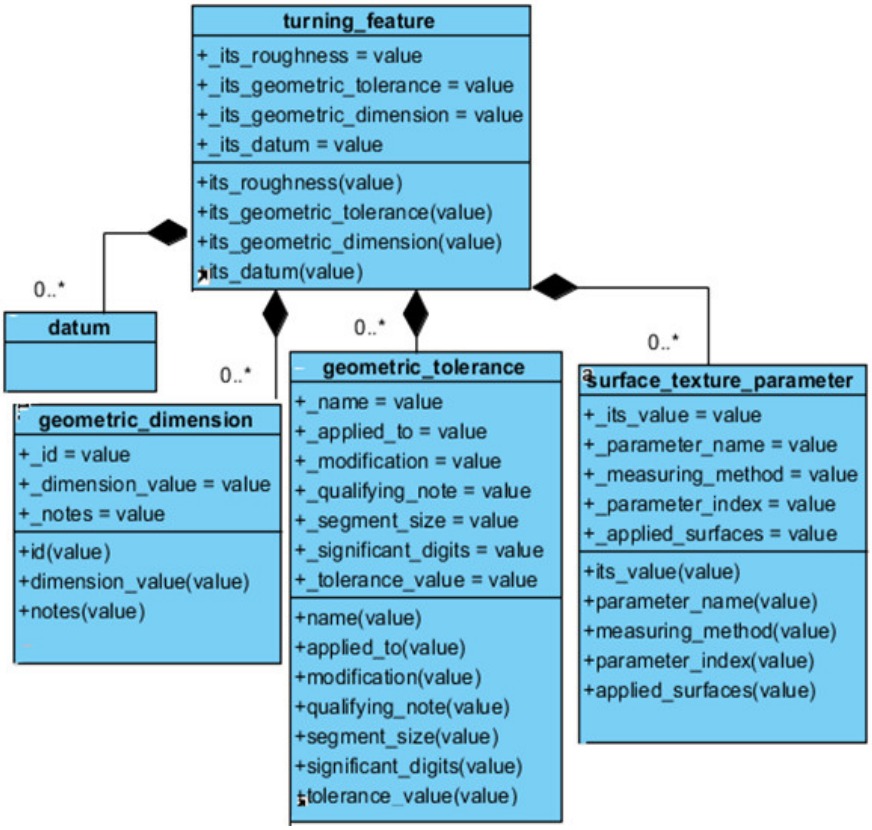


Figure 5: New association between turning feature and related tolerances

After describing how surface roughness is embedded in the STEP AP238 data file and how tolerances can be associated to the related manufacturing features,the parameters representing tolerances will be searched and located using defined methods.

Figure 6 show the class outer_diameter containing the concerned tolerances inherited from the turning_featureparent class.

It must be noticed that surface_texture_parameter has been taken just an example for its simplicity; the new inherited classes can also handle other types of tolerances defined in STEP AP238 such as, geometric tolerances, geometric dimension and Datum.

The input file that has been taken in this example will just consider the roughness since it is the only tolerance that has been applied to this feature. Other tolerances types won't be instantiated.

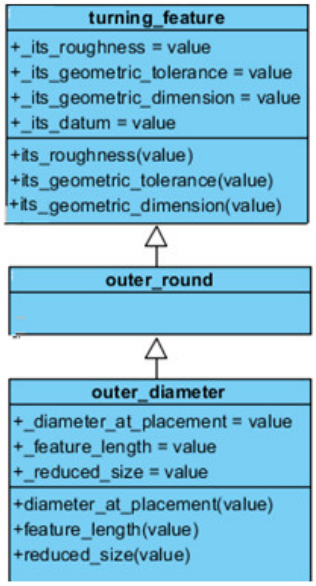


Figure 6: The new association representing the outer_diameter class that inherit the tolerances type

Therefore, the main tasks in GD&T extraction and features/tolerances association are to extract, instantiate then associate between each manufacturing feature and its related tolerance as an attribute.

Figure 7 illustrates the flowchart of the algorithm established for the surface_texture_parameter extraction from STEP AP238, and its association with the corresponding feature.

For the positional roughness, the algorithm is divided into a three-stage process.

In the first stage, the callout object and the positional SURFACE_TEXTURE_PARAMETERare located.

The second stage consists in instantiating a new object of type surface_texture_parameter. The last stage consists in associating the surface_texture_parameter to the corresponding manufacturing feature.

Taking the fact that the ADVANCED_FACEs that compose each feature are known, the new tolerances objects will be then associated to the corresponding feature by applying the new association method explained in this work.

That solution will lead us to get objects inherited from turning_feature, representing tolerances related to the corresponding manufacturing features.

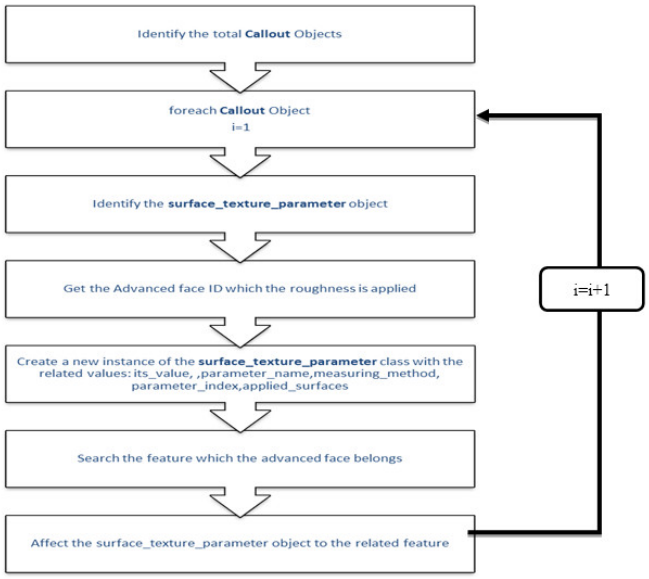


Figure 7: Flowchart of the association between the surface_texture_parameter object and its related manufacturing feature

5 Conclusion

In this paper, we have developed a new approach to extract tolerances from a step ap238 file, and then associate them with their related machining features. An algorithm was implemented and embedded into an expert system called ICAPP-TURNto extract and map tolerances data for process planning. The main task is to locate each tolerance and associate it with the related feature.

The data integration module described in this paper contains the GD&T extraction methods from a step ap238 data input file. These methods locate the surface roughness, its value and the advanced face to which the roughness is applied using custom defined functions. The data representation module creates the mapping of the related machining features and tolerances to the oriented objects, then makes associations between these objects based on a new approach for conception and implementation of object oriented that improves the reusability , flexibility and the relationships between the machining features and their tolerances as properties. These modules represent an input data for ICAPP-TURN system.

Thus, in a future work, we will describe the knowledge representation module which provides the knowledge to the centralized database. This module receives the input knowledge about rules and constraints, and then represents this information in a form that ICAAP-Turn can utilize to store data and resolve tasks.

REFERENCES

[1] Antonio Armillotta,A method for computer-aided specification of geometric tolerances,Dipartimento di Meccanica, Politecnico di Milano, ViaLa Masa 1, 20156 Milano, Italy, Computer-Aided Design 45 (2013) 1604–1616.

[2] Mehmet I. Sarigecili , Utpal Roya,Interpreting the semantics of GD&T specifications of a product for tolerance analysis, Computer-Aided Design 47 (2014) 72–84.

[3] OntoSTEP Plug-in. July, 2012. http://www.nist.gov/el/msid/ontostep.cfm.

[4] ISO_10303-11, Industrial automation systems and integration — Product data representation and exchange —Part 11: Description methods: The EXPRESS language.

[5] ISO_10303-21, Industrial automation systems andintegration — Product data representationand exchange —Part 21: Implementation methods: Clear textencoding of the exchange structure.

[6] Pei Lei, Lianyu Zheng,A STEP-NC compatible on-machine measurement system with automated correlation of inspection data. Applied Mechanics and Materials Vols. 397-400 (2013) pp 1772-1777

[7] HakkiEskicioglu, The Use of Expert System Building Tools in Process Planning, Middle East Technical University, Ankara, 1992, Engineering Applications of Artificial Intelligence.Vol. 5, No. 1, pp. 33-42, 1992.

[8] JaiderOussama, al. Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, Issue 5(Version 1), May 2014

[9] STEP Application Protocol (AP) 203 editions 1 & 2,Configuration Controlled 3D Designs of Mechanical Partsand Assemblies, ISO10303 1994, (2007) & (2009).

[10] http://step-nc-machine.software.informer.com/9.1/

[11] http://www.steptools.com/

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Migration of the Temporal RDB into Temporal ORDB including Bitemporal Data : Phases

[1]Ain El Hayat Soumiya, [2]Bahaj Mohamed
[1]LITEN Laboratory, Faculty Of Science and Tecnology,University Hassan , SETTAT, Morocco
[2]PROFESSOR, LITEN Laboratory, Faculty Of Science and Tecnology,University Hassan , SETTAT, Morocco
soumya.ainelhayat@gmail.com, mohamedbahaj@gmail.com

ABSTRACT

This paper proposes an approach for migrating existing relational database (TRDB) according to SQL: 2011 standard into temporal object relational database (TORDB) including Bitemporal data. This is done with methods that can extract various functions from a TRDB, which is based on the different kind of the relationships between tables. To do that we're going to enhance a representation of a varying time database's structure in order to make hidden semantic explicit. In contrast to other studies, our main goal here is to offer a first and better solution to mentioned limits to existing works, in order to provide the efficient method for the translation from TRDB to TORDB. We are going to take an existing RDB with bitemporal data features as input, we provide our meta-model in order to get a correct description of a data structure, enriches its metadata representation, and generates a bitemporal data Model (BTDM), which captures essential characteristics of temporal databases for migration. A prototype has been developed which proves the effectiveness of this solution, and temporal queries has been made with help of Oracle 12C.

Keywords: Bitemporal Data, SQL:2011, Temporal database, RDB,ORDB, Semantic Entichment

1 Introduction

Temporal database is one of the most important parts of the information technology. It is generally known that temporal database stores the history of the objects or the database activity. Even today, a large number of database applications are based on time in nature. It is an important attribute of each and every real world application. Every event important for operation of enterprise must be record at specific point at time. Events occur at specific points in time, objects and the relationships among objects exist overtime [1]. This importance made it necessary to make better description and clearly some tasks of database system. Especially, the need to retain trace and audit the change made to a data and the ability to plan based on past or future assumptions are important uses cases for temporal data [2]. Many database applications require management of time varying data. The most common example of such applications involves health care management, medical records of patients, reservation system, banking and accounting, and decision support system.

The literature on temporal database offers three dimensions of time for temporal data support, which are independent each to other: transaction time, valid time, user-defined time. Valid time is the period during which a row is occurred in the reality in the database, it is the fact as is valid in the modeled in the real.

DOI: 10.14738/tmlai.54.2977
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2977

Transaction time automatically captures changes made to the state of time-variant data in a database [3].User-defined time is a time representation implemented to satisfy specific needs of users. There is also another dimension of temporal database called bitemporal data. Bitemporal database systems support both valid time and transaction time. On the other hand, bitemporal databases model our changing knowledge of the changing world, hence associate data values with facts and also specify when the facts were valid. There by providing a complete history of data values and their changes [4]. Several temporal data models, which support different dimension of time, are discussed in previous work.

The relational database model has come a long way since the 1970s and it has become the dominant model of database [5], most traditional database applications are based on traditional management systems. Although, the relational database is designed to serve many applications which need only store the recent state of data, they are insufficient for those which need to retrieve past as well as current and future [6] .The need to shift time manipulation from application to relational database is identified and implemented in SQL: 2011 standard. One of the advantages of the SQL: 2011 aspect, based on period time. The temporal relational database with SQL: 2011 features have been accepted as a solution for kept data changes over time with insertion of new records or update old records. However, many problems have been emerged, the weakness of such Temporal RDBMSs in supporting complex data structures, user-defined data types and data persistence required by temporal object relational database. Furthermore, the reconstruction of temporal complexes objects split across relational tables is costly because its causes many join. ORDBs with time-varying features has addressed of these problems, which showing potential, because they have a relational base and append object concept, to enhance ORDB and give the correct description of records , we associate time at attribute. Temporal ORDB overcomes the disadvantage of data redundancy introduced when using temporal RDB to store the information.

The purpose of this paper is to propose a method for migrating bitemporal data from RDB based on SQL: 2011 standard, into Temporal ORDB integrating time-varying features. The method comprises three basic phases. In the first, we will examine the useful features of temporal data in order to create an RDB meta-Model, which holds the necessary elements for the correct description of temporal database. In the second step, the method takes the entire bitemporal relational database and stores it in a temporal structured table that contains several parameters, attributes, class, relationships, cardinalities, integrity constraint and time periods, in order to enrich and realize the schema translation (BTDM). The BTDM so obtained is converted into a temporal ORDB model, which treats complexes object and data semantic that can expressed in its metadata in the third step. The prototype has been developed to demonstrate the migration process, and TORDB queries example will be generated to valid our approach.

2 Related Work

Significant researches address the problem of numerous temporal models and query language. Atay compared interval-based attribute and tuple time stamped bitemporal data models, accesses and evaluated their usability using the same data and same queries for both approaches. According to this comparison, Petkoviç examined the performance implication for tuple and attribute timestamping, her test stored data using two different forms, and perform the 36 query on both.

A research work in [7] proposed a temporal object relational SQL language handling valid time at the attribute level in a temporal transparency environment his paper. The approach in [8] presented a

database application development environment for both temporal and non-temporal using SQL: 2003 following the attribute timestamping. Comparison of three different storage models (OODB, ORDB, and XML) for the parametric temporal data model and estimate storage costs are discussed in [9].

Oracle added temporal processing to its database technologies. There are a few work deals with integrating temporal dimension to an existing DBMS. Sandro Radovanović evaluated performance of traditional relational DBMS (RDBMS) and temporal databases using Oracle 12c DBMS [10]. Verification of temporal data using valid time dimension support in Oracle is proposed in [11]. Ptrovic uses the most important temporal concepts to investigate their implementations in enterprise database systems such as Oracle, DB2 and teradata[12].

The ISO (international organization for standard) and IEC (International Electrotechnical Commission) committee, initiated o project to create a language extension to support temporal database, is given in [13]. The most important features in SQL: 2011 to create and manipulate temporal database implemented by IBMDB2, is discussed in [14].

From the all-overhead results, we conclude that most proposed solutions contain limited and simple rules compared to our work, notably regarding the scope and methodology. We believe that studies are based on time-variant data schema to make records of data, not even completed to provide a comprehensive data model, especially at the representation level. Furthermore, Some semantics aspects are not considered in the previous works .Our study is using the concept of semantic enrichment that give the possibility to understand the structure , meaning of temporal databases and construct schema translation which is enhanced by additional data semantic.

During our criticized analysis, we feel that some aspects of time variant data was overlooked in many works, which reflect that the challenge for those authors was only to cover the technical part of the storage, retrieval rather than on migration and gain from the offered advantages in temporal object relational, in order to reduce the redundancy of data by supporting the attribute timestamping.

The goal of our work is not mainly to create better temporal objet relational database from BTDM, but to afford a well arranged and a complete as possible transformation from temporal RDB into temporal ORDB, that can be a concrete reference for further investigations and works in this common area. we summarize and illustrates clearly the completeness of our mapping strategy, that include entities, objects, annotation, relationships between classes, attributes in their various forms, data types, value types, class constructs, constraint types, valid time period, Transaction period and much more.

We create our solution by combine several results from the existing methods and apply our enhancement using some semantics concepts. More precisely, we propose the rules that facilitate the transformation from Temporal RDB into temporal database based on ORDB using bitemporal data dimension. Therefore, this study presents a meta-model to define set of stereotype for the specification of new characterization of the bitemporal data associated with UML class diagrams, which simplify the creation of BTDM.

3 Process of Migration from TRDB into TORDB:

In this section, we outline the important phase for translation. In the fist, we provide the TRDB design, and then we will define the BTDM and TORDB model.

3.1 Creation of TRDB design :

The main purpose behind constructing a TRDB design is to simplify the comprehension of essential metadata stored in temporal databases. An efficient TRDB modeling overcomes the complications that occur during matching period time and keys in order to classify relations between classes, attribute, relationships and dimensions of time. An important advantage of TRDB design is that it identifies the migration of periods, therefore adding more semantics to a TRDBs metadata.

Consider the database shown in Fig1, which modeling the purchase orders administration. This model will be used in the examples presented along the paper. Primary keys are underlined in bold ex, and Foreign Keys are marked by "*".

-Customer (**IDcust**,name,city,street,phone,*OrderNo ,**vt-start,vt-end**,TT-start,TT-end)
- Company (***IDcust**,Type, taxes)
-Customer-Ass (**IDcust**,descpt,percent, **vt-start,vt-end**,TT-start,TT-end)
- Purchase-Order (**OrderNo**,tocity,tostreet,tozip,shippingdate, *lineNo, **vt-start,vt-end** TT-start,TT-end)
- OrderLine-Item (**lineNo**,*ProdNo , Quantity, **vt-start,vt-end**,TT-start,TT-end)
- Products (**ProdNo**,descpt,price, **vt-start,vt-end**,TT-start,TT-end)
- Store (**IDstore**, location,capacity,street,city)
- Stock (***ProdNo** , ***IDstore**, Quantity, **vt-start,vt-end**,TT-start,TT-end)

Figure1.Sample input temporal Relational database

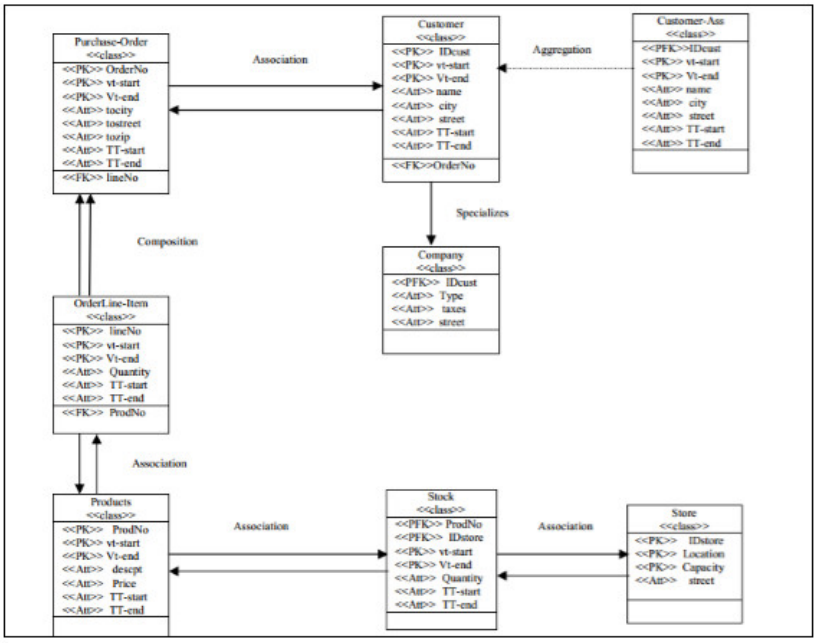


Figure 2: TRDB Meta-Model

The creation of TRDB design is the first step of the migration into TORDB. To do this, we define a new set of UML model elements that provides mechanisms that enable new kinds of modeling elements to be defined, and relate the information to new modeling elements. This is accomplished by integrating

stereotype, constraints and tagged values.Fig2, illustrates the meta-model that is proposed for relational databases, which support SQL: 2011 standard.

3.2 Semantic enrichment of temporal relational Database: BTDM

Semantic Enrichment is a process of analyzing databases to understand their structure, and to make hidden semantics explicit [15].To enrich the semantic of temporal RDB schema, we have to extract its data semantic to be enriched and converted into a much enhanced BTDM. To do this task, the process starts by extracting the basic information about an existing temporal RDB schema, integrating relations names, Periods time, and Attributes properties. We assume that data dependencies are represented by primary keys and foreign keys, as for each foreign key value, there is a reference to an existing PK value. The next step is to identify the BTDM constructs based on classification of data and the relationships, which may be performed through data access. In the last step, the BTDM structure is generated.

3.3 Definition of BTMD

The BTDM is a representation of RDB using SQL: 2011 Standard, which is enriched with semantic data in order to provide a new type of tables describing the different classes, extracted from temporal RDB with the data necessary for the creation of temporal ORDB. This phase produces a data reference model that is designed to allow the exchange the schema and the sharing of information to reuse.

The BTDM is defined in our approach as a set of element:

$$BTDM := \{ C \mid C := \{Cn, Att, Rel, clas, bitDim\}\}$$

Where:

- Each class has a name Cn
- Att= denotes a set of Attributes of class C
- Att={a|a:=(N,T,L,NL,D,tag)} , where N is an attribute name , T is an attribute type, L: data length, NL: if the attribute accepts the parameter null or not(N/ NoN),Tag: primary key (PK)| foreign key(FK)| primary foreign key (PFK).
- REL: Relations BTDM each class C has a set of relationships with other classes ,where rel is defined in C with another class C' REL={rel | rel:= RelType,DirC,Car } ,where RelType denotes a relationship, DirC is the name of C' that interacts with C, Car means cardinalities describing the relationship.

  RelType supports five types of relationships, and accepts the following values:

  1. "Ass" for Association
  2. "Agg" for Aggregation
  3. "comp" for Composition
  4. "inher" and "inherBy"for Inheritance

- Clas: classification

  Classification divides classes into two different kinds of categories:

  1. Temporal class (TCls): class contains a varying time period. The time period during which a row is regarded as correctly reflecting reality by the user of database [16].

2. Simple class or No Temporal class (SCls): class without varying time attributes.

- BitDim: Bitemporal data dimension

  Each temporal class has a valid time period, transaction period or the both. BitDim is used to specify time variant attributes:

  BitDim={B | B=: {VT, TT}}

where:

VT: is valid time period, it can have the following values:

VT={VT_LB,VT_UB}, which specify the lower and upper bound of the valid time data value.

TT: is transaction time period, TT accepts two values:

TT ={TT_LB, TT_UB}, which specify also the lower and upper bound of the transaction time.

3.4 Generation of BTMD

In this paper, the BTMD is considered the interesting phase for the migration process which in the end generates the target scheme.

Cn	Clas	Attribute					Rel			BitDim			
		N	T	L	NL	Tag	RelType	DirC	Car	VTL	VTU	TTL	TTU
Customer	TCLS	IDcust	Number		NoN	PK	Ass	Purchase-Order	1				
		name	Varchar	25	NoN		Ihne by	Company	1				
		city	Varchar	25	NoN		Agg	Customer-ass	1...N				
		street	Varchar	25	NoN					Y	Y	Y	Y
		phone	Number		N								
		OrderNo	Number		NoN	FK							
		vt-start	Date		NoN	PK							
		Vt-end	Date		NoN	PK							
		TT-start	sysdate		NoN								
		TT-end	date		NoN								
Company	TCLS	IDcust	Number		NoN	PFK	Ihne	Customer	1				
		Type	Varchar	25	NoN					Y	Y	Y	Y
		taxes	Varchar	25	NoN								
Customer-Ass	TCLS	IDcust	Number		NoN	PFK	Agg	Customer	1				
		Descpt	Varchar	25	N								
		percent	Varchar	25	NoN								
		vt-start	Date		NoN	PK				Y	Y	Y	Y
		vt-end	Date		N	PK							
		TT-start	sysdate		NoN								

		TT-end	date		N								
Purchase -Order	TCLS	OrderNo tocity tostreet tozip lineNo vt-start Vt-end TT-start TT-end	Number Varchar Varchar Varchar Number Date date sysdate date	 25 25 25	NoN NoN NoN NoN NoN NoN NoN NoN NoN	PK FK PK PK 	Ass Com	Customer OrderLine-Item	1…N 1	N	N	Y	Y
OrderLin e-Item	TCLS	lineNo ProdNo Quantity vt-start vt-end TT-start TT-end	Number Number Number Date Date sysdate date		NoN NoN NoN NoN N NoN N	PK FK PK PK 	Com Ass	Purchase-Order products	1..N 1..N	Y	Y	Y	Y
Products		ProdNo descpt price vtstart vt-end TT-start TT-end	Number varchar Number Date Date sysdate date		NoN NoN NoN NoN N NoN N	PK PK PK 	Ass Ass	OrderLine-Item Stock	1 1..N	Y	Y	Y	Y
Store	SCLS	IDstore Location Capacity street city	Number Varchar Number Varchar Varchar	 25 25 25	NoN NoN NoN NoN NoN	PK 	Ass	Stock	1..N	N	N	N	N
Stock	TCLS	ProdNo IDstore Quantity vt-start Vt-end TT-start TT-end	Number Number Number Date Date sysdate date		NoN NoN NoN NoN NoN	PFK PFK PK PK	Ass Ass	Products Store	1 1	Y	Y	Y	Y

Figure 3: Result of BTDM Generation

3.5 Definition and Identification of TORDB Model:

This step, presents the different elements composed a TORDB Model, which provide a complete description of temporal OR database. The TORDB model is defined as a set of temporal typed table based on temporal structured type TST for storing data.

Each ST consists of a set of non varying-time attributes and varying-time attributes defined as primitive collection data type or reference pointing to a specific ST or TST. In this paper, the varying time attribute can be a bitemporal attribute, which actually stored in nested table collection type. Each value assigned to the temporal attribute must respect the temporal data representation of a temporal column.

Definition of TORDB model: is denotes as three-tuples:

TORDB Model = {TTs, STs ,Tm}

Where: TTs is set of temporal and non temporal typed table, STs is a set of temporal structured type or simple structured type , and Tm is a time-varying Period. The sets TTs, STs and Tm are defined as follows:

- STs ={Sn , S, AT},where Sn is the name of a structured type , S is the super type of ST, and AT is a set of structured type's attributes:
  AT={A | A:={N,T,D,NL,BitT,M}},where N: is the name of attribute, T: means data type which can be primitive , UDT or reference. N: if the attribute accepts Null or not. D: default value. M: denotes if the AT is a single valued or collection valued. BitT: denotes if the attribute contains a bitemporal attribute is defined:
  BitT={(AT1,AT2,….,VTL,VTU,TTL,TTU)}.
- TTs ={typedtable|Ttable={TTn,STn,PK,Tp}}where TTn is the name of typed table, STn is the name of the structured type based upon which TT is defined, PK : primary key, TP: means if the TT is temporal or not.

3.6 Creating tables:

It can be able to produce temporal ORDB queries for relationships with oracle 12C which is formed by the temporal and no temporal queries. We use Oracle's concept of nested table to create the varying-time attributes. Example1 show the creation of the customer table and of all necessary auxiliary (object) types.

Example1:

```
Create type Bitemporal_T as object{
TT_LB date,
TT-UB date,
VT_LB date,
VT_UB date}/
Create type Bit_period is table of  Bitemporal_T;
Create type CustAss-T as object(
Type varchar(25),
Taxe varchar(25))
Create type PurchOrd_T as object {
```

```
OrderNo Number,
Tocity varchar(25),
Tostreet varchar(25),
Tozip varchar(25),
history  Bit_period}/
Create table PurchaseOrd_T of PurchOrd_T CONSTRAINT OrderNo-PK PRIMARY KEY(OrderNo), NESTED TABLE  Purch-history
STORE AS Bitemporal_tab;
Create type NT_PurchOrd as Object{
OrderNo Number,
VT_Start date,
VT_End date}/
Create type PurchaseOrder_Hist  is table of  NT_PurchOrd;
Create type Customer_T as Object{
IDcust Number,
Name varchar(25),
city varchar(25),
street varchar(25),
phone Number,
Customer_Association   CustAss-T,
Cust_history  Bit_period,
Purchase_Order  PurchaseOrder_Hist
}/
Create table Customer of Customer_T CONSTRAINT Cust-PK PRIMARY KEY(IDcust), NESTED TABLE Cust_history  STORE AS
Cust_tab, NESTED TABLE  Purchase_Order  STORE AS  PurchOrd_tab;
```

Figure4: TORDB Queries: Example

4 Conclusion

This work outlines the basics phases of migrating from RDB according to SQL: 2011 standard into ORDB including bitemporal data, with a simple and practical method to capture the relationships between different kinds of classes. Currently, no approach has proposed such a solution to extract data model from RDB implemented by SQL: 2011 aspect. This approach is superior to existing work as it generates the ORDB with bitemporal data, including the schema and data semantics, and it exploits the range of powerful features provided by SQL standard.

Our method exceeds the existing works as it generates the TORDB tables by creating a BTDM from a TRDB, and we use it as an input enriched with semantic data, and this last provides a TORDB Model to capture the characteristics of temporal and non temporal SQL query.

A forthcoming paper will propose an algorithm for converting method from TRDB into TORDB, with Bitemporal data that not requires any human interference.

REFERENCES

[1] Jaypalsinh A. Gohil, Dr. Prashant M. Dolia ."Checking and Verifying Temporal Data Validity using Valid Time Temporal Dimension and Queries In Oracle 12c", International Journal of Database Management Systems (IJDMS), Vol.7, No.4, August 2015.

[2] M. Kaufmann, P. M. Fischer, N. May, D. Kossmann: "Benchmarking Bitemporal Database Systems: Ready for the Future or Stuck in the Past?" . EDBT, 738-749, 2014

[3] D. Petković , "Performance Issues Concerning Storage of Time-Variant Data, Egyptian Computer Science Journal", Vol. 38 No, 2014.

[4] Atay, C. , " A Comparison of Attribute and Tuple Time Stamped Bitemporal Relational Data Models", Proc. of the Int. Conf. on Applied Computer Science, p: 479-489, 2010.

[5] D. Brdjanin , S Maric , Z S Pavkovic «On Suitability of Standard UML Notation for Relational Database Schema Representation » . Enterprise Business-Process and Information Systems Modeling , Vol 248, pp 399-413, 2016

[6] Dušan Petkovic. "Temporal Data in Relational Database Systems: A Comparison". New Advances in Information Systems and Technologies,Vol 444,pp 13-23. 2016

[7] Chau, V.T.N. and Chittayasothorn, S. "A Temporal Object Relational SQL Language with Attribute Timestamping in a Temporal Transparency Environment,"Data & Knowledge Engineering, (67), p. 331-361, 2008.

[8] T. Chau and S. Chittayasothorn, "A Temporal Compatible Object Relational Database System", The IEEE Southeastcon, 2007.

[9] Noh, SY., Gadia, S.K. & Jang, H. J. Cent, "Comparisons of three data storage models in parametric temporal databases". Journal of central south university , vol 20: 1919_1927, 2013.

[10] Slavimir Vesić, Slađan Babarogić Nenad Aničić, "Use of the Temporal Concepts in Transaction Database", PROC. SYMORG, p 850-857 , 2014.

[11] ISO/IEC 9075-2:2011, Information technology-Database languages-SQL - Part 2: Foundation (SQL/Foundation), 2011.

[12] Kulkarni, K., Michels, J. "Temporal Features in SQL: 2011", SIGMOD Records, Vol. 41, No. 3, 2012.

[13] Maatuk, A., Ali, M. A. and Rossiter, N.: "Semantic Enrichment: The First Phase of Relational Database Migration". In CIS2E '08, 6pp, Bridgeport, USA, 2008.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

IoT Middleware Architecture based on Ontologies to Model Logistic Process

S. Aoulad Allouch, K. Amechnoue, I. Achatbi
National School of Applied Sciences Tangier
saloua.aoulad@gmail.com; kamechnoue@gmail.com;i.achatbi@gmail.com

ABSTRACT

In recent years, the Internet of Things (IoT) become a promising topic of technical social and economic significance, especially with the high number of developed sensors and technologies. Logistic applications are a perfect domain of IoT as it adds new functionalities in identification, traceability, storage and real-time tracking of good in the supply chain. Handling the huge quantities of heterogeneous IoT components and logistic items is an important challenge. We present in this paper a new Internet of Things middleware architecture for logistic transport applications. This architecture focuses mainly on a semantic model that uses ontologies for sensor data representation by describing the main entities involved in the logistic scenario.

Keywords-component; IoT; Middleware; semantic and ontology; supply chain; logistic

1 Introduction

IoT can be traced back to the pioneering work done by Kevin Ashton from Massachusetts Institute of Technology (MIT), Auto-ID Center in 1999, and was initially linked to the new idea of using radio frequency identification (RFID) and electronic product code (EPC) in the supply chain [1]. It is envisioned that billions of objects will be equipped with different kinds of sensors, actuators and mobile devices and connected to the internet via heterogeneous access networks.

Logistics is the art of managing, controlling and implementing the flow of goods, services and related information between the point of origin and the point of consumption. In the other hand, logistic has become an important component of economy and more interest are given to this domain. As it is necessary to have professional and intelligent logistical support to achieve international trading and distribution process, IoT is mainly applied. In fact, IoT will provide intelligence to the logistics distribution network, transparency and real-time to the items management of the logistic system, moreover it can reduce the storage cost and improving the logistics service quality, so that the whole supply chain integrates more closely.

A logistic domain may have a relatively large number of applications and information sources, likewise the IoT components are heterogeneous, the number of sensor and actuators involved are growing, so a massive real-time data flow will be produced by this devices. The data provided by different objects, sensing devices and that relating to the logistics domain, must be organized and interpreted in a

DOI: 10.14738/tmlai.54.2978
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2978

homogeneous way to allow interoperability in the logistics chain, likewise define corresponding events for the different situations of products.

In the light of the challenge above, we propose an IoT middleware architecture based on semantic and ontology to describe the sensing devices and their functionalities in IoT, thus, to model the logistic domain and facilitate collaboration and interoperability between system components.

The remainder of the paper is organized as follows: Section 2 gives an overview of the background of our work from the vision of middleware, semantic and ontologies. The related work concerning logistic and sensor ontology is briefed in Section 3. Section 4 presents the proposed IoT middleware architecture and its main elements. Finally, in section 5 we give a conclusion and we discuss briefly our future work.

2 Background

In the context of logistics applications, our goal is to enhance collaboration between stakeholders in logistic chain for an intelligent monitoring, and a better coordination and visibility to respond to customer orders and market demand. This by given more attention to traceability and real-time tracking of goods in the chain. IoT can act as the basis for the development of services and applications that offer information sharing between logistics partners, this, provide improved visibility of supply chain. To model logistic process with the IoT, the integration of several technologies is very important and it involves three major categories: Radio-Frequency Identification (RFID), sensors and embedded systems for identifying, sensing and reading logistic items. Due to the large variety of used technologies, IoT requires interoperability between devices and information, to address this concern; we propose an IoT middleware for logistic applications based on semantic and ontology.

2.1 Middleware

The middleware is an interface that facilitates the interaction between technological and the application levels, this with providing technical interoperability. IoT is characterized by a heterogeneous and dynamic infrastructure comprised of an ultra large number of things, so, collecting data from these different objects is an important task as it allows software systems to understand the environment better. Further, we cannot expect all these objects to be connected to the computers due to technical and economic reasons. In this context middleware has a major role in simplifying the development of new services and the integration of legacy technologies into new ones, thus, IoT middleware sits between the IoT hardware and data and the applications that developers create to exploit the IoT.

2.2 Semantic Web and Ontologies

The term “Semantic Web” refers to World Wide Web Consortium (W3C) vision for the future of the Web in which information is given well defined meaning, making it simpler for machine to perform more work involved in finding, combining and acting upon information on the web. Thus, enabling computers and people to work in cooperation. The collection of Semantic Web technologies (RDF, OWL, SKOS, SPARQL, etc.) provides an environment for people to create data stores on the web, build vocabularies, and write rules for handling data.

Since the early 1990s, ontologies have become an important research topic examined by numerous research communities, several definitions of ontologies have been suggested in the last decades. But the one that best describes ontologies can be found in Gruber [2] “ontology is a formal, explicit specification

of a shared conceptualization". Similarly, the W3C[3] defines ontology as the terms used to describe and represent an area of knowledge, and is used by people, databases, and applications that need to share domain information. Ontology has been widely used to facilitate knowledge sharing and reusing. For pervasive computing environments, using ontology to model context and situation enables multiple entities to have common understanding of the context and situation during collaboration. An otology comprises three elements: First, classes in the many domains of interest, second, relationships that can exist among things, third, the properties those things may have.

3 Related work

3.1 Sensor ontology

Nowadays, most research and standardization related to IoT domain have principally concentrated in sensor description and observation data modeling. Thus, researchers have realized that the data sharing and service interoperability between different data models and devices can be handled using semantic web technologies. In addition; they confirm that ontology serves as a key factor to enable interoperability between heterogeneous systems. Several ontologies have been presented for sensor. The work in [4] presents an attempt to capture the most important features of a sensor node that describes its functionality and its current state. The ontology describes the main components of a sensor node such as processor CPU, power supply, and radio and sensor modules. OntoSensor[5] constructs an ontology comprised of definitions of concepts and properties adopted in part from SensorML[6]an extensions to the IEEE Suggested Upper Merged Ontology (Sumo)[7]. However, it does not provide a descriptive model for observation and measurement data.

The W3C Incubator Group on Semantic Sensor Networks (SSN) has introduced ontology[8] to describe sensors and sensor networks. The ontology represents a high-level schema model to describe sensor devices, their capabilities, platform and other related attributes in the semantic sensor networks and the sensor Web applications. Nevertheless, SSN ontology does not include units of measurement, location, hierarchies of sensor types and domain knowledge that are related to sensor data. Thus, concepts were included to allow linking to such external ontologies.

Semantic Sensor Observation Service ontology (SemSOS or O&M-OWL) [9] models sensor observations and measurement in OWL, it provides the ability to query high-level knowledge of the environment as well as low-level raw sensor data, besides it adds semantics essentially to weather domain.[10]has proposed an ontology that models three aspects of the real world present in the IoT; things, real world concepts and functionalities of things and real world approximation, and this by linking, the domain of knowledge for sensing, actuating, and processing tasks and the real world representation through IoT services that are aware of their environment. The principal idea of this ontology is that it creates a direct relation between physical concept, mathematical formulas and functions.

3.2 Logistic ontology

Many initiatives have been taken by researcher in the domain of logistic ontology:

[11]Proposed a logistic ontology based on situations semantic by identifying a set of logistic events and logistics situations founded on logistic process that model the transition of product. Thus, they use a situation model based on semantic to specify a situation of product and determine events caused by this situation. [12]Contribute in the advancement of logistic ontologies by referring to supply chain

management and reusing the existing body of knowledge contained in the SCOR (Supply Chain Operation Reference) model. As the SCOR model does not aim at covering interorganizational logistics to the full extend, this ontologies lack some characteristic concepts of this domain.

In another tentative,[13]build a logistic domain ontology model (LDOM) that represent relations among logistic domain knowledge. This ontology contains 12 top classes – namely cargo, organization, supplier, customer, carrier, transport service, constraint, transport service standard, transport mode, vehicle, traffic line and information between actors. [14]Proposes a core ontology that determines the principal concepts usually adopted in the logistics domain, thus it can be used as the basis to promote communication, facilitate integration of heterogeneous IT system between organizations, and finally it is used for engineering purposes.

4 Global Overview of the Proposed Architecture

4.1 Global IoT architecture for logistics

The proposed IoT logistic architecture presented in Fig, adopts a service oriented approach as generally done in the literature, based especially on semantics to describe devices, their data, and their physical attributes. The adoption of the SOA principles allows for decomposing complex and monolithic systems into applications consisting of an ecosystem of simpler and well-defined components [15][16],thus, it is divided into four layer: perception layer, network layer, semantic middleware layer and application layer as designed in Figure 1.

Perception layer:

The focal mission of this layer is to perceive the physical properties of things during the logistic process. This information is about state of good and vehicles or their environment such as: temperature, humidity, orientation, chemical changes in the air… etc. This process of perception is based on several sensing technologies (e.g. Sensor, RFID, NFC, GPS, etc.).Hereafter, the collected information are converted to digital signal and then passed to Network layer for their transmission.

Network layer:

This layer is mainly responsible for the transmission of information collected in perception layer to middleware layer. The transmission medium can be wired or wireless and technology can be 3G, UMTS, Wi-Fi, ZigBee, Bluetooth etc. depending upon the sensor devices.

Semantic Middleware layer:

This layer sits between applications, and the network layer and helps brings together a multitude of devices and data in a way that permits developers to create a new IoT services without having to know the detail of each device or data format.

Our future IoT middleware is an extended of the work done in [17], it involves four important parts: a discovery module, service management, service composition and a knowledge base.

Discovery module consists of two parts: device discovery which enables any device in the IoT network to detect all its neighboring devices and make its presence known to each neighbor in the network. Service discovery: is used to find appropriate devices according to the service description submitted by the requestor.

Service management provides the main functions that are expected to be available for each objects and that allow for their management in the IoT scenario.

Service composition gives the functionalities for the composition of single service offered by networked objects to build specific applications; in this component we find the process responsible for coordinating all tasks.

Knowledge Base is the main component of this middleware; it contains all information needed by other piece, and gives a comprehensive set of ontologies that describe logistic items, logistic process, sensor, actuator and all physical concept and unit of measurement needed in our case.

Application layer:

The principal goal of this layer is the management of application based on the objects information handled in the middleware layer, in our case it comprise all functionalities offered to stakeholders in logistic process, as tracking and checking the state of goods and environment, react to changes in task and network requirements, thus, sharing important information with the right actor at the right time, and responding dynamically to the order market demand.

In our work, more attention will be given to the semantic middleware layer especially on modeling a set of ontologies to describe goods, environment and the devices used in the logistic process.

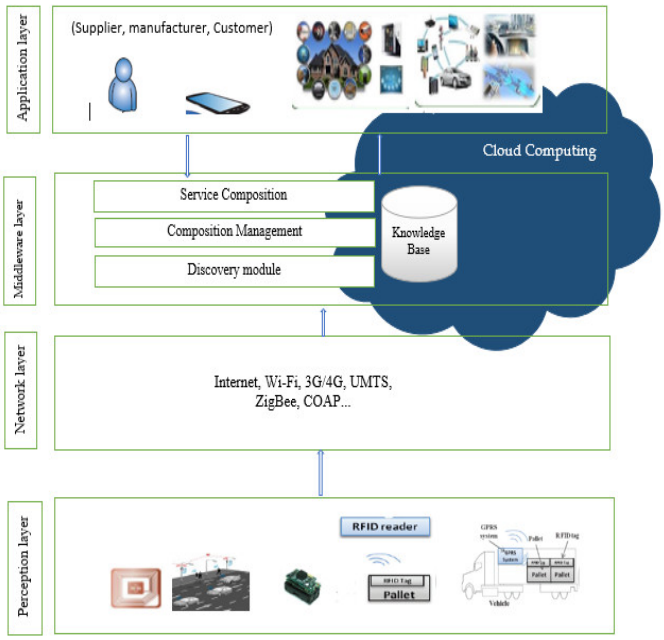


Figure 1: General IoT middleware Architecture for logistic process

4.2 Proposed ontologies

In this section, we will present a set of ontologies that models device, thing and logistic. Our work is based on the advantages and shortcomings of the ontologies listed above and by combining the work done in [10], thus adapting to our case study that concern logistic transport.

4.3 Device ontology

Device ontology describes functionalities, hardware attribute and physical concepts measured by devices: sensor, tags, and actuator. Based on work done in[10][18]and according to the requirement of logistic process we have detected the principles ontology concepts used to better model devices. A representation of the main class related to the device is shown in Figure 2.

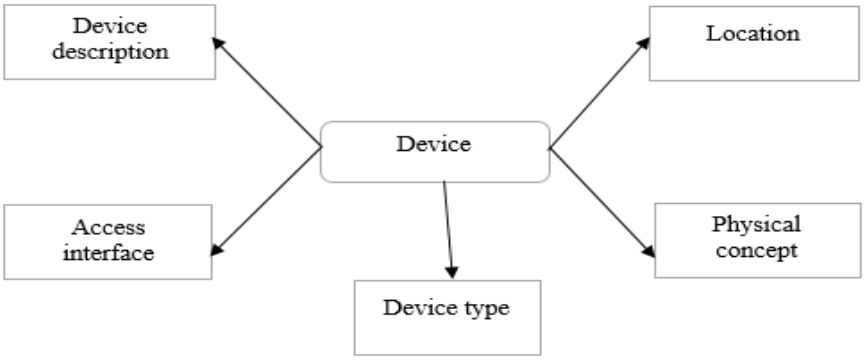


Figure 2: Device and related first class entities

Device type: device can be divided in three main type: sensor, tag, or actuator. When the type is sensor, our ontology will be connected to an external ontology SSN.

Device description: it contains information about device's manufacturer, device's internal hardware components and all information that can be specific to each device, which give a good vision about the quality of device hardware.

Physical concept: it contains real world properties measured by device such as: humidity, temperature, pressure…etc. Thus their value and unit of measurement.

Location: defines the location of the device and it is connected to an external ontology: local and global location ontologies

Access interface: it includes technologies generally used in distributed systems such as: RPC, SOAP, and REST.

4.4 Thing ontology

The goods transported during the logistic process have some specific features that need to be taken into consideration, for this reason we have dedicated an ontology especially to model information related to product: their state, location, manufacturer, condition of conservation…etc. The main classes linked to thing concept are grouped in Figure 3.

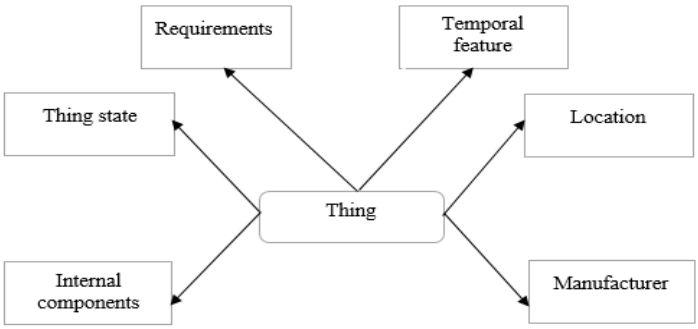


Figure 3: Thing and related first class entities

In fact, manufacturer class contains the historic of product from his point of origin and including all transitional activities to integrate product's component. Internal components regroup the raw materials used for the production of product. Moreover, requirements involve all information needed to transport product in a good conditions such as: high and low temperature allowed, incompatible product, dangerousness constituents. Location tracks the emplacement exact of product and it is similar to the one described in device ontology. Thing state comprises all observable features that we can need to know the state of product such as: temperature, pressure, humidity…etc. Temporal feature specifies the chronologic of product and the way in which his properties are transformed.

4.5 Logistic ontology

This ontology is based on core ontology proposed in [14] and it will be extended according to our use case scenario. The most relevant objects for logistic domain are: transport, kind of good, mode of transport, type of equipment used. In addition to these concepts, other logistic situation and events are discussed in [11]and they will be the basis for our ontology. We will describe in more detail in future work this ontology in a logistic transport situation and by using 4PL provider.

5 Conclusion

We presented in this paper general middleware architecture for logistic process, based on semantic technologies, especially a set of ontologies that provide interoperability between heterogeneous components and models the different elements of IoT involved in the domain of logistic. The relevant aspects are presented by three ontologies: device ontology to describe sensing device and their functionalities, Thing ontology for product and their requirement, and logistic ontology that show most concepts used in the logistic domain.

Our purpose in the future work is to implement the proposed ontologies and evaluate their feasibility, thus handle other part of middleware to improve communication and keep services loosely coupled in order to increase reusability.

REFERENCES

[1] Z. Sheng, S. Yang, Y. Yu, A. Vasilakos, J. Mccann, et K. Leung, « A survey on the ietf protocol suite for the internet of things: Standards, challenges, and opportunities », IEEE Wirel. Commun., vol. 20, no 6, p. 91–98, 2013.

[2] T. Gruber, « What is an Ontology ».

[3] « OWL Web Ontology Language Overview ». https://www.w3.org/TR/owl-features/.

[4] S. Avancha, C. Patel, et A. Joshi, « Ontology-Driven Adaptive Sensor Networks. », in MobiQuitous, 2004, vol. 4, p. 194–202.

[5] D. J. Russomanno, C. Kothari, et O. Thomas, « Sensor ontologies: from shallow to deep models », in System Theory, 2005. SSST'05. Proceedings of the Thirty-Seventh Southeastern Symposium on, 2005, p. 107–112.

[6] «Sensor Model Language (SensorML) | OGC ». http://www.opengeospatial.org/standards/sensorml.

[7] «The Suggested Upper Merged Ontology (SUMO) - Ontology Portal ». http://www.adampease.org/OP/.

[8] « Incubator Report - Semantic Sensor Network Incubator Group ». https://www.w3.org/2005/Incubator/ssn/wiki/Incubator_Report.

[9] C. A. Henson, J. K. Pschorr, A. P. Sheth, et K. Thirunarayan, « SemSOS: Semantic sensor observation service », in Collaborative Technologies and Systems, 2009. CTS'09. International Symposium on, 2009, p. 44–53.

[10] S. Hachem, T. Teixeira, et V. Issarny, « Ontologies for the internet of things », in Proceedings of the 8th Middleware Doctoral Symposium, 2011, p. 3.

[11] P. Lian, D.-W. Park, et H.-C. Kwon, « Design of logistics ontology for semantic representing of situation in logistics », in Digital Media and its Application in Museum & Heritages, Second Workshop on, 2007, p. 432–437.

[12] J. Leukel et S. Kirn, « A supply chain management approach to logistics ontologies in information systems », in International Conference on Business Information Systems, 2008, p. 95–105.

[13] X. K. Zhang et T. Tian, « Logistics domain ontology model and its application », in Advanced Materials Research, 2010, vol. 108, p. 1403–1408.

[14] L. Daniele et L. Ferreira Pires, « An ontological approach to logistics », 2013.

[15] T. Teixeira, S. Hachem, V. Issarny, et N. Georgantas, « Service oriented middleware for the internet of things: a perspective », in European Conference on a Service-Based Internet, 2011, p. 220–229.

[16] S. Bandyopadhyay, M. Sengupta, S. Maiti, et S. Dutta, « Role of middleware for internet of things: A study », Int. J. Comput. Sci. Eng. Surv., vol. 2, no 3, p. 94–105, 2011.

[17] S. Hachem, « Service-Oriented middleware for the large-scale mobile Internet of Things », Université de Versailles-Saint Quentin en Yvelines, 2014.

[18] S. De, P. Barnaghi, M. Bauer, et S. Meissner, « Service modelling for the Internet of Things », in Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on, 2011, p. 949–955.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Hello Message Scheme Enhancement in MANET based on Neighbor Mobility

Berradi Halim, Habbani Ahmed
SIME Lab, ENSIAS, University of Mohammed V, Rabat, Morocco
berradi@ifride.com, habbani@ifride.com

ABSTRACT

The key idea behind the routing protocols based in multi hope is to find the required path from the source node to the destination. Since those protocols does not consider the node mobility in their mechanism, we propose enhanced the routing protocols by benefit from localization node and predict the time needed for neighbors to go out of range for a node to increase the robustness of the protocols with the mobility. First, we propose an efficient scheme to predict the neighbor rang out. Second, we designed an enhanced version of OLSR (Optimized Link State Routing) based on the collected data. Third, we conduct an extensive set of simulations to compare the performance of our proposal against OLSR origin.

Keywords: component; MANETS, Network Cartography, Node Localization, Position prediction, OLSR Protocol, Routing Validity.

1 Introduction

Due to MANET (Mobile ad hoc network) features, Mobility is the main challenge to build efficient routing protocols. Even though some protocols like OLSR[1] protocol is conceived specially for these networks, it is still unable to completely fit their inherent characteristics. In fact, node mobility significantly degrades its functioning.

According to the OLSRstandard version, the only mean to better track the timely changing network topology consists in properly tuning the periodicity of the control messages. However, the scarcity of the wireless network resources prevents the viability of such a solution. To adapt OLSR to the characteristics of such dynamic networks, several extensions and improvements of the basic OLSR were proposed, such as the F-OLSR [2](Fish eye OLSR), P-OLSR[3] (Position-based OLSR).

There is two classical routing strategies used in proactive protocols, such as link-state routing (e.g., OLSR) or distance-vector routing (e.g., DSDV[4]). The major disadvantage of these approaches is the consumption of the available bandwidth if the topology change frequently and the network suffer from frequent path breaks.

In this work, we propose to utilize the Node position as a ground basis on which we perform our routing decisions to overcome the mobility effect. The remainder of the paper is organized as follows. Section II presents some relevant related work. In section III we talk about the impact of the mobility in the routing protocols. In section IV, we detail our proposed prediction model. Section V is devoted to define the

DOI: 10.14738/tmlai.54.2979
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2979

performance metrics used to evaluate our scheme against OLSR based on extensive simulation results. Finally, we conclude the paper in section VI.

2 Related works and motivation

In [5]They propose a protocol which improves OLSR by taking into account the node mobility and signal strength in the selection of MPRs. And in the papers of [6] and [7] they propose that the node in wireless Ad Hoc network knows its position to achieve the precise positions of every node in the network. In [8]They utilize the network cartography in different ways in order to improve the efficiency of the routing function. In [9]They propose the integration of a cartography gathering scheme to enhance the capacity of the Optimized Link State Routing Protocol (OLSR) to properly track node movements in dynamic networks. In [3]They discuss problems of OLSR or other routing protocols that are due to mobility of nodes. They propose enhanced OLSR protocol (P-OLSR) in packet delivery, throughput, and latency and normalized overheads. In this work [10] , they propose to increase the network lifetime using prediction of residual time to select the stable MPR. In our knowledge, no one use node position to improve OLSR protocol against mobility, thus we will work in this enhancement.

3 Impact of the mobility on the network

In this section, we first define the simulation setup, to measure the impact of the mobility on the standard protocols between three protocols (OLSR, AODV and DSDV) in the MANET in term of the lost packet, delay and the overhead. Then, we conduct a series of simulations in NS3 (Network simulation 3), and finally, we interpret the result.

3.1 Simulation setup

We established a network consists of 30 nodes in the network simulator NS3 [reference], we conducted several experiments were distributed to 25 test during 200 seconds. To generate traffic in the network, 10 nodes are randomly selected to be a source of CBR (Constant Bit Rate) traffic. And these selected nodes use UDP (User Datagram Protocol).

Note that since the nodes are mobile and we are in random simulation environment, we repeat every simulation 25 times to achieve a good simulation results. The entire node moving randomly using “RandomWayPointMobility” in the simulation.

In the TABLE.1 below, we show our simulation parameters used during simulations:

3.2 Results and discussions

To show the impact of the mobility on the three selected protocols (OLSE, AODV, DSDV), we used two significant metrics, which we consider important in mobile network, lost packet and the delay sum.

Lost packet

In Figure 1 we plot the lost packet over speed between the three protocols, we observe that due to the mobility of the nodes, the lost packet increase when the speed increasing.

The AODV had more lost packet than the DSDV, and the OLSR had the less lost packet than the other protocols. And this is a logically result, because the AODV is a reactive protocol, so he will suffer from the break links more the two other protocols.

Parameters	Values
Modulation	802.11b
Nodes	30
Mobility Model	RandomWayPointMobility
Simulation time	200 (s)
Packet size	256 (bytes)
Protocols	OLSR, AODV, DSDV
Speed	[0,10,20,30,40,50,60,70,80,90,100]
Simulation Area	5000*5000

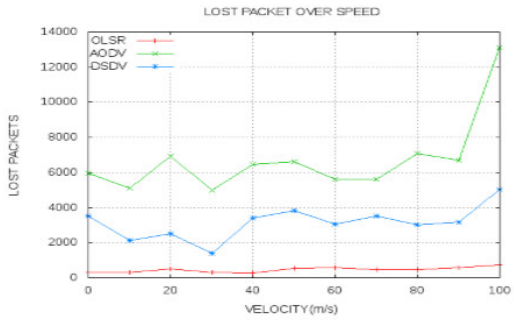


Figure 1 the lost packet over speed

The Delay Sum

We get the delay sum by calculate the cumulative of the delay of all the flow between the source and the destination in the simulation. In Figure 2 we see the same observation as the lost packet, AODV have the greatest delay sum, and OLSR have the smallest one.

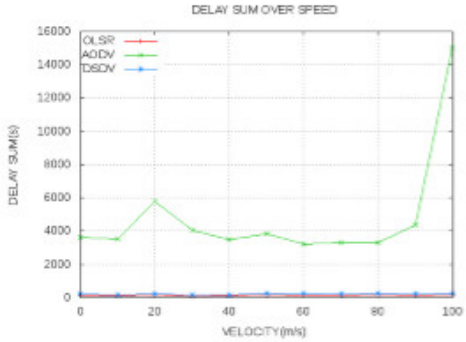


Figure2 the delay sum over speed

Thus, OLSR is the best protocol against the mobility between the three selected protocols. In our proposal, we will make OLSR operate better than the standard protocol, by adding intelligent in send hello periodic time.

4 Our contribution

Our proposal

OLSR is a proactive routing protocol for MANET designed to operate in distributed areas and it is well suited for dense and mobile networks. It has the advantage compared by other classical protocols such as

the stability, minimizes flooding traffic, reducing the retransmit control messages. The basic functions of OLSRare: detection of links and neighborhoods and selection of the MPRsin order to disseminate the topology information to other nodes. Detection of links is achieved by transmitted a periodic HELLO messages between the nodes, generated at static time interval (every 2 seconds in the standard). The drawback of this specification is that we do not learn about the mobility of the network. Because the more the networks move, the more we need to update the topology. Our proposal algorithm (Figure 3) is to changethe constant HELLO periodic time (2 seconds)to a variable that depend on the network mobility.

Every node will calculate the remaining time for every neighborhood to quit from his wife range, and we will take the smallest value as a periodic time.

So, the periodic time will depend on the network mobility, the more network move the more nodes send message hello, the more network move less the less nodes send message hello.

When any nodes go out of the range of node A, the node must send the next message hello to update the topology. That why we choose the smallest time in buffer as a HelloTimeIntervale. The remaining time will be $\in]0,\infty]$. Therefore, to conserve an optimum running for the protocol we will limit to the interval $\in]0.5,5]$

Algorithm: // at reception of any message HELLO

In this algorithm

R: wife range of the node

Buffer: an array of the estimated time for every neighborhood

HelloTimeIntervale: is the hello time-intervals (2 seconds in the standard)

T: is the necessary time for a neighborhood to get out of the range of A

Begin

1. Get the position Xa and Va velocity from the current node
2. Get the position Xb and Vb velocity of the sender node (from the message hello)
3. Calculate the estimated time T for the node B to go out the range R of node A (see Figure 4)

If (T > 5)

Than T←5

Else if T<0.5

Than T←0.5

4. Save the time in the Buffer
5. HelloTimeIntervale← min value from the buffer

End

Figure 3 Algorithm used to calculate the necessary time for a neighborhood to get out of range of another node

Prediction time model

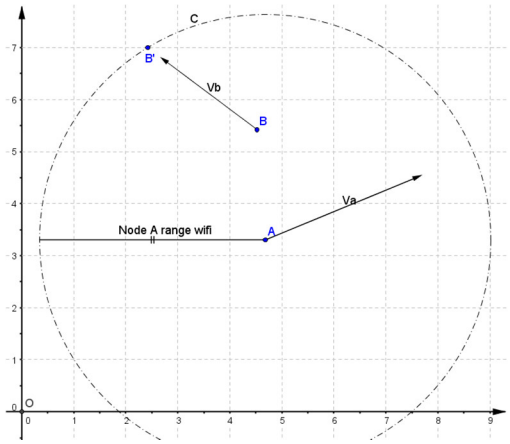


Figure 2 A sample on Predict Remaining Time Estimation

As showing in Figure 4, our proposal is to estimate the necessary time τ for Node B to move to Position B' (we are in two-dimensional coordinates), so:

$$\tau = \Im(V_A, V_B) \tag{1}$$

We know that:

$$\overrightarrow{V_A} = \begin{cases} V_{A_x} = \dfrac{dx_A}{dt} \\ V_{A_y} = \dfrac{dy_A}{dt} \end{cases} \longrightarrow \begin{cases} x_A(t) = V_{A_x} \times t + x_{A_0} \\ y_A(t) = V_{A_y} \times t + y_{A_0} \\ A_0(x_{A_0}, y_{A_0}) \quad when \quad t = 0 \end{cases} \tag{2}$$

$$\overrightarrow{V_B} = \begin{cases} V_{B_x} = \dfrac{dx_B}{dt} \\ V_{B_y} = \dfrac{dy_B}{dt} \end{cases} \longrightarrow \begin{cases} x_B(t) = V_{B_x} \times t + x_{B_0} \\ y_B(t) = V_{B_y} \times t + y_{B_0} \\ B_0(x_{B_0}, y_{B_0}) \quad when \quad t = 0 \end{cases} \tag{3}$$

And

Point B' is the intersection between the Node A range Wi-Fi and Node B.

Therefore, the Node B' verifies the circle equation (C):

$$(x_{B'} - x_A)^2 + (y_{B'} - y_A)^2 = R^2 \tag{4}$$

We replace x_A, x_B, y_A and y_B with the corresponding equation, we will get a finale equation:

$$at^2 + bt + c = 0 \tag{5}$$

With:

$$\begin{cases} a = (V_{B_x} - V_{A_x})^2 + (V_{B_y} - V_{A_y})^2 \\ b = 2((V_{B_x} - V_{A_x})(x_{B_0} + x_{A_0}) + (V_{B_y} - V_{A_y})(y_{B_0} + y_{A_0})) \\ c = (x_{B_0} + x_{A_0})^2 + (y_{B_0} + y_{A_0})^2 - R^2 \end{cases} \tag{6}$$

Since our network move randomly, so $a >= 0$

Therefore, the resolut ion of this equation is:

$$\text{if } a>0 \begin{cases} t_1 = \dfrac{-b+\sqrt{\Delta}}{2a} \\ t_2 = \dfrac{-b-\sqrt{\Delta}}{2a} \\ with \quad \Delta = b^2 - 4ac \end{cases} \qquad \text{if } a=0 \begin{cases} t = \dfrac{-c}{b} \end{cases} \tag{7}$$

In general, we have two solutions and we will choose the positive one, because the prediction time is in the future so it will positive (when the node A receive Hello message $t_0 = 0$).

5 Implementations

Modified OLSR

The OLSR protocol make eachnode periodically sends HELLO messages to its neighbors to broadcast information about their own existence by using the techniques of link detection and the neighbor discovery. Thus, in order to enable the node to get the position and the velocity information of the neighbors, the format for the HELLO message is modified. Six fields are added which is showed inFigure 5.Thecoordinates and the velocity of the node (xPos, yPos, Vx, Vy).

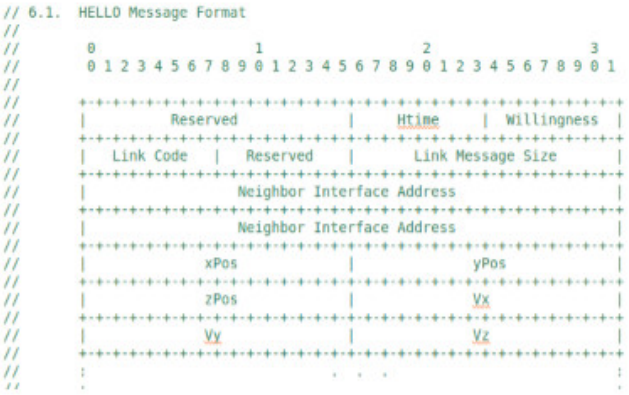


Figure3. The modified HELLO message format

These coordinates can be used to calculate the estimated time to go out of range of the node that will receive the HELLO message.

After receiving the HELLO message from the neighbor nodes, link sensing is performed by the node, and the local link information base is created and saved. As for link sensing, we will execute our algorithm in Figure.3 to set the new HelloTimeIntervale and see the performance of our proposal.

Propagation radius of the Node

Toward running our simulation, we must fix the nodes propagation radius R in our algorithm, this radius is about 350m, therefore when we put a 500m between two nodes and run our simulation. We observe that no HELLO message was received. We justify this by the noise and the size of HELLO message. In order to verify the distanceradius that conserves a communication between two nodes (with the HELLO messages).

We changed thedistance between the twonodesto observethe impact on the exchange HELLO message. The following experimental environment is created. The scenario consists of 10 nodes arranged in a grid (4*3). The distance between the nodes is changed between 0 and 400. Nodes transmit periodically packets containing 250 bytes. The simulation run executes for 200s. Figure.6 is the distribution for the nodesin the grid. The experimental results indicated (see Figure.7) that after the distance between 2 nodes be 350m, we node can receive any HELLO messages. As a result, we will choose the radius R = 350 meter in our algorithm (see Figure.3).

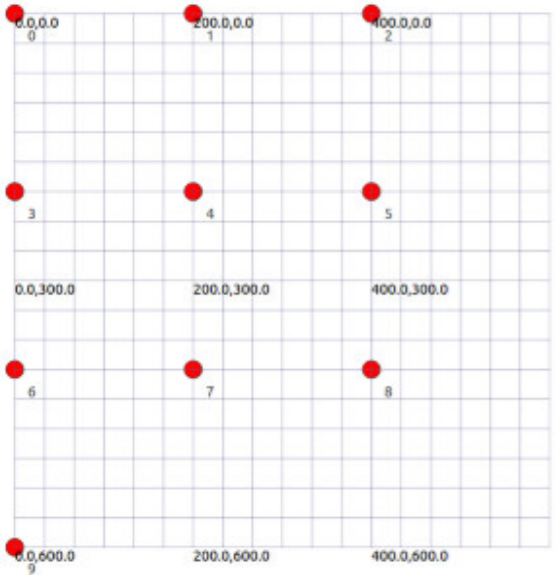


Figure 6. The distribution for the nodes in the grid

messages. As a result, we will choose the radius R = 350 meter in our algorithm (see Figure.3).

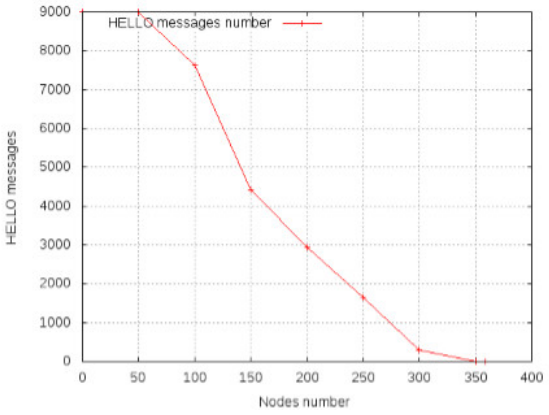


Figure 7. The impact of distance on the HELLO messages exchange

6 Comparison and discussions

In this section, we describe our results of our proposal in compare with the standard version of OLSR. Specifically, we compare the lost packet and the delay sum.

We set the parameters of the simulation as the first experiment settings (see Tab1). Thus, the number of nodes is set to 50, whereas 10 randomly nodes send the data packet of 256 bytes and the others received.

The node speed is changed between 0m/s and 100 m/s. The wireless range Ris fixed to 350 meters (experimentally) to assume that there isanexchange of HELLO messages their geometric distance is smaller than the wireless range. As mobility models, we use the random waypoint.

Lost packet

The following figure (Figure 8) represents the number of lostpacket inOLSR protocol and in Modified-OLSR protocol for different speeds. This figure shows that for different values of velocity, the modified OLSR protocol has less lost packet than the standard version. We can consider that the modified OLSR shows an enhancementwhich may can't be negligee. This extends considerably the quality of the services (QoS) in a mobile network.

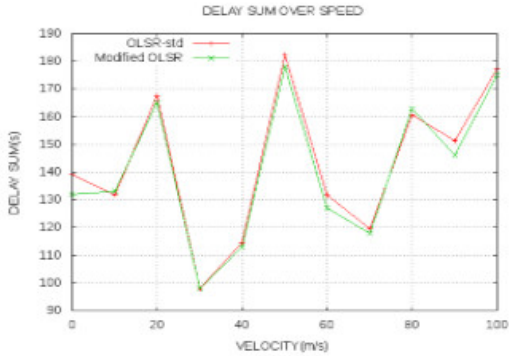
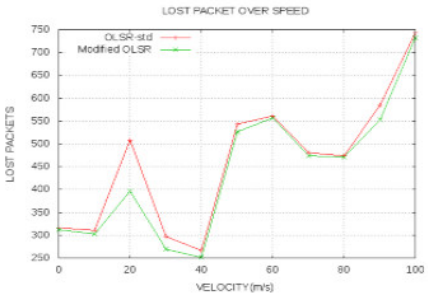


Figure 7 the delay sum over speed

The delay sum

We get the delay sum by calculate the cumulative of the delay of all the flow between the source and the destination in the simulation. In Figure 2 we see the same observation as the lost packet, AODV have the greatest



delay sum, and OLSR have the smallest one.

Figure 9 The lost packet over speed

In the second figure (Figure 9), we have a slightly difference between the Modified OLSR and the OLSR standard, but the modified show better delay than the standard version. Thus, Location information can be utilized to assist in decreasing the transmission delay. The simulation studies demonstrate the impact of using geographical information to enhance the QOS in high-speed mobile nodes.

7 Conclusion

The node localization technology is one of the important technical in applied researches for the wireless Ad Hoc network. The goal of this paper was to analyze the impact of mobility on routing protocols of ad hoc networks, and proposes an enhancement of the protocol OLSR based in geographical position of the neighborhood. For this purpose, we aimed to extend the lifetime of paths used for routing the data traffic and the control messages by reducing the effect of node mobility. So, we predict remaining lifetime of neighborhood based on distance and we send the HELLO message when a node quite the wife range to update the topology information. Absolutely, the simulations have showed that PDR and DELAY SUM are improved compared to the standard version of the OLSR.

Although, this proposal was developed for OLSR protocol, the same technique can be applied in different protocols. As part of our future work, we plan to studies the impact of this technique in AODV protocol, since it suffers the most from the mobility of the nodes.

REFERENCE

[1] Jacquet, P., et al. Optimized link state routing protocol for ad hoc networks. in Multi Topic Conference, 2001. IEEE INMIC 2001. Technology for the 21st Century. Proceedings. IEEE International. 2001.

[2] Adjih, C., et al., Fish eye OLSR scaling properties. Communications and Networks, Journal of, 2004. 6(4): p. 343-351.

[3] Sharma, S. P-OLSR: Position-based optimized link state routing for mobile ad hoc networks. in Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on. 2009.

[4] Perkins,C.E., and Bhagwat.P. (1994). Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. ACM, pp.234-244.

[5] Dashbyamba, N., et al. An improvement of OLSR using fuzzy logic based MPR selection. in Network Operations and Management Symposium (APNOMS), 2013 15th Asia-Pacific. 2013.

[6] Wang, A. and B. Zhu. Realize localization in routing table in wireless networks. in Computer Science and Network Technology (ICCSNT), 2012 2nd International Conference on. 2012.

[7] Wang, A. and B. Zhu. Realize 1-hop node localization based on OLSR protocol in Ad Hoc networks. in Computer Science and Network Technology (ICCSNT), 2012 2nd International Conference on. 2012.

[8] Belghith, A. and M. Belhassen. Routing enhancements in dynamic MANETs with obstacles. in Wireless and Mobile Networking Conference (WMNC), 2011 4th Joint IFIP. 2011.

[9] Belhassen, M., A. Belghith, and M.A. Abid. Performance evaluation of a cartography enhanced OLSR for mobile multi-hop ad hoc networks. in Wireless Advanced (WiAd), 2011. 2011.

[10] Ouacha, A., et al. Reliable MPR selection based on link lifetime-prediction method. in Networking, Sensing and Control (ICNSC), 2013 10th IEEE International Conference on. 2013.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Contribution to the Measurement of Organizational Performance based on A Multi-Agent Approach

Ansar Daghouri, Khalifa Mansouri, Mohammed Qbadou
Laboratory : Signals, Distributed Systems and Artificial Intelligence (SSDIA)
ENSET of Mohammedia, Université Hassan II of Casablanca
dagh.ansar@gmail.com, khmansouri@hotmail.com, qbmedn7@gmail.com

ABSTRACT

This research focuses on evaluating and analyzing the organizational performance of a risk management unit within banks. The main proposal is to analyze and simulate the process of risk management based on decision support system and artificial intelligence. This is why this paper uses the systemic thinking and simulation tool. We finally propose a multi-agent model showing nine autonomous agents communicating with each other to simulate a risk. This model provides both a tool to simulate the risk and a way to modify the organizational structure of the risk management unit to improve the performance of bank.

Keywords : Organizational performance; Decision Support Systems; Artificial Intelligence; Systemic thinking.

1 Introduction

The business of banking is the ability to manage risks. These risks are unexpectedly recurrent in terms of their number and frequency. Good risk management is affected by the structure and the internal organization of the unit itself. This work presents an approach which aims at building a decision support system dedicated to the analysis of the organizational performance of a risk management unit. It is based on the systemic approach through the following steps: the observation to delimit the system and its interactions, the static modeling to describe the system in a formal way according to certain standards, the identification of the indicators in order to focus on the weak-spots and the strength of the unit, and the dynamic modeling by using multi-agent system in order to simulate the technical subtleties of the studied system.

2 Position of the problem

2.1 Bank risks

Risks refer to a well-identified threat associated to the occurrence of an event or series of events which are perfectly descriptive, and it is not known whether they will occur once again or not but they are still likely to occur in an exposing situation[1].

The bank like any other organization faces risky conditions. The risks which a bankmay face are variable and both their nature and extent are relative. However, in addition to the diversity of types of risks and the narrow or extended scope of each type, they are classified as follows:

DOI: 10.14738/tmlai.54.2980
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2980

2.2 Typology of bank risks [2]:

Credit risk

It is the risk in which financial losses are due to the clients' or counterparties' inability to honor their financial commitments. In other words, it is the risk that the borrower will not repay his debt at a certain deadline.

Market risk

Market risk is when the bank undergoes financial losses as a result of changes in stock prices, currency or interest rates.

Operational risk

It is linked to the risk of loss which results from maladaptation or imputable failure due to internal procedures, personnel and systems or external events.

Risk of non-conformity

It is the risk of penalty or financial loss which arises from non-compliance with the financial regulations of banks or their internal rules.

Liquidity risk

This is a risk for which the bank may not be able to meet its obligations as they fall due. It results from a bank's inability to meet a reduction in its liabilities or to finance an increase in its assets.

Interest rate risk

Interest rate risk is identified as the negative impact of the adverse changes in interest rates on the financial position of the bank. Interest rate risk concerns both the interest rate positions taken in trading halls and the exposure to thetransformation risk that is inherent in banking by definition.

Strategic risk

It is a risk which is inherent in the chosen strategy or which results from the inability to implement this strategy. A strategic failure can be very serious because the committed resources become worthless and hence the losses become very significant.

Reputation risk

It arises from a negative perception by customers, counterparties, shareholders or investors, which may adversely affect the ability of the bank to maintain business relationships and the continuity of access to sources of finances.

2.3 Risk management

Risk management in banks is a topical issue from which many countries suffer (United States, Japan, Great Britain, France…). This phenomenon is the result of economic problems.

A few years ago, Morocco also got close to falling into financial instability.Thanks to the government's initiative, banks were able to avoid this crisis before it could create more dangerous impacts.

These financial crises, which have attacked the Moroccan market in particular and the international one on a large scale, have clearly highlighted the weaknesses in risk management in the banking sector. [3]

Traditional management based on simply the compliance with the prudential rules has proved ineffective because of lacking the involvement of management and the board of directors in banking organizationcontrol.

Under these conditions, it becomes clear that a global approach to banking risk management is no longer sufficient, hence the need for the involvement of management and the board of directors in choosing the most appropriate tools .

Banks Vs. risk management

Banking risk management corresponds to all the techniques and tools put forward by banks to measure and monitor the risks which they may face.

The banking institutions are engaged with different degrees of involvement in the implementation of management tools allowing early monitoring of the different types of risks.

The leaders must have procedures and tools for managing the various risks; this management brings together three fundamental elements:

Risk assessment: it is a continuous process with three stages namely the rigorous analysis and the quantification of risks, the determination of the level of risk tolerated by the bank, and the comparison of the level of risk tolerance with the scale of its assessment;

Risk management and control: following the assessment of risks and the level of risk tolerance, the bank's management must take steps to manage and control the risks such as the implementation of policies and the security measures ;

Continuous risk monitoring: this is the most important step in the management process since banking activities are rapidly evolving at the rate of innovation. They use new technological tools such as the Internet hence the need for tests and audits (internal and external).

2.4 Performance and measurement tools

Performance

The term of performance is widely used in the field of management sciences, according to Bourguignon , in spite of the fact that there is no single and unanimous definition of this term. Its etymological origins signify the accomplishment of a task in order to evoke afterwards the exploit and the success.

Based on the state of the art, performance is defined as the combination of effectiveness, which is the ability to achieve the objectives and the efficiency which is the relationship between the means and the efforts being made.

Performance can not be defined as a simple concept because it is a notion that encompasses several dimensions. In this paper, we are interested in organizational performance which is defined as the measures which are directly related to the organizational structure and not on its possible social or economic consequences.

In fact, the performance of an organization is measured by performance criteria; there is no single criterion or an exhaustive list covering the various performance criteria.

Method and tool for measuring performance

«Performance measurement describes and implements the strategy, the monitoring of organization, decision-making and improving coordination» [4].

The state of the art has not made it possible to distinguish a single and a common method for the measurement of organizational performance.

However to construct the indicators for performance measurement, there are different currents of thought[5].

These different currents make it possible to construct the organizational indicators;the main task of an indicator is to measure the situation to trigger the decision-making.

3 Modeling and simulation

A risk is a complex situation that needs to be managed. Banks are subject to undergo stress at this kind of event and must make decisions conveniently. One of the identified means is the use of modeling and simulation to analyze this problem.

The systemic [6] represents a particular approach which is adapted to understand the behavior of a complex system in order to model it to act and to simulate its variations in time.

A complex system, by definition, is a system that consists of many components, interconnections, interactions or interdependencies which are difficult to describe, understand, predict, manage, conceive, and change[7].

3.1 The systemic approach

The systemic approach is a means to understand the system and the organizational complexity. Its stages highlight two levels: the system which is constructed by an observer facing a complex situation [8] and the model which is the representation of a system. The purpose of its construction is to understand the degree of importance of the decision-making on the modeled system.

In this paper, we have relied on the work of Donnadieu and Karsky [9] who explain the systemic approach in five stages: observation, systemic exploration, qualitative modeling, dynamic modeling and simulation.

Several methods of modeling of processes [10] exist in literature, namely: the dynamics of the systems [11], the SADT [11], the FIS method[12] and the UML formalization[13].

The choice is based on the UML language: it describes the system , its functions, the resources, the hierarchy and schematically the process.

The UML approach will be used only to formalize the interactions within the system. For the dynamic modeling, we are based on the decision support systems (DSS).

Multi-agent systems (MAS)[14] constitute another way to implement tools for the decision-making. They consist in managing the agents by providing them with tasks for cooperation.

3.2 Multi-agent systems (MAS)

MAS are entities that interact with one another to produce a collective conduct. One of the main characteristics is the distribution of tasks on several agents that are objects capable of interacting and performing actions in parallel with other agents.

The agents must be coordinated with one another, which would allow to model the systems, the modes of communication and the simulations in the field. The interaction is essentially through the transfer of information which is broken down into three parts: receiving information, reasoning on other agents and issuing the messages or the actions to be performed.To build a MAS, there are several platforms[15]: Jade, MadKit, Zeus and Jack.

4 Dss and the evaluation of organizational performance

4.1 Work methodology

The proposed methodology to analyze the performance of banks' risk management unit is based on a systemic approach. This approach allowed us to analyze the system and model it to reduce its complexity. The proposed methodology provided a risk presentation to analyze the bank's weaknesses. This approach is based on five main axes:

Observation, to know the organization of the risk management unit and to identify its actors as well as the interactions between them;

Systemic exploration, which allows modeling the system studied and the different interactions between its components and the environment. For our case, modeling is done by the UML language;

The identification of performanceindicators, which will subsequently be integrated into the dynamic model for the analysis of the risk management unit;

Dynamic modeling, which adds the notion of time to the qualitative model. We have worked with MAS that have reduced the complexity of problem solving by dividing the tasks on agents;

Bank Evaluation: it allows having a general overview of the studied system in order to improve the performance of the bank.

This approach will make it possible to study, to model and to simulate the tasks of the risk unit to provide at the end a vision of a well-formalized decision support.

4.2 The risk management unit

The risk management approach in banks goes in parallel with the professional and the regulatory standards whose management rules are defined according to international levels which are recommended by the authorities. The Risk Management Unit is independent of business lines and professions and it reports directly to the Presidency. Its main task is to supervise and hedge the various risks which are inherent in the activities of the bank, and also to control and measure everything.

4.3 Risk modeling

To understand the interaction between the actors in a risk management process, we propose a generic use case diagram(Fig.1):

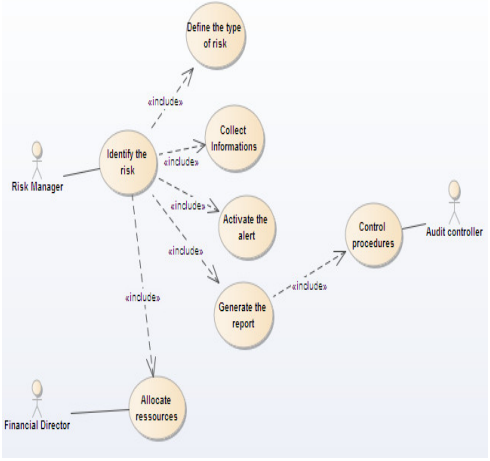


Figure.1. Risk Use Case Diagram

The risk management process begins when the Risk Manager identifies the risk, then it goes on to define the type of risk we have previously classified, the collection of information required for processing,internal rules and procedures, the triggering of the alert according to the type and severity of the risk, the preparation of a risk management report, which will be subsequently monitored by the audit controller, the request for the availability of financial resources from the financial management.

4.4 Modeling of the risk management unit

After identifying the actors of the unit and their roles, we have modeled the unit with the class diagram of the UML language.The unit consists of a set of committees, each of which is characterized by resources and interacting with each other.

We have worked on the most general case to arrive at a global bank model that is represented by the class diagram in Fig. 2.

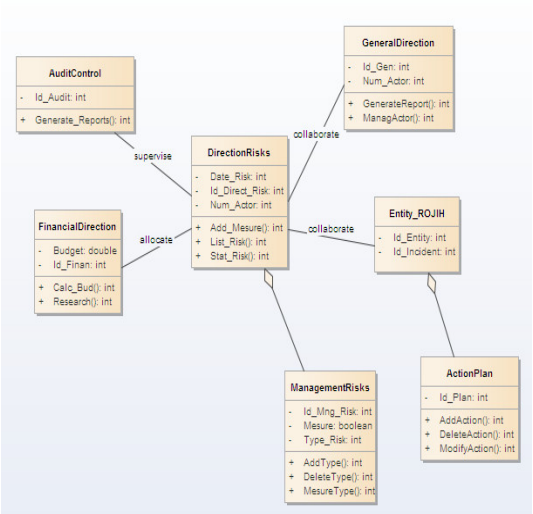


Figure.2: Unit class diagram

To model the risk management unit, it has been necessary to present all the departments that cooperate with the latter in order to better manage all the risks a bank can face. The class diagram consists of seven classes:

GeneralDirection : its objective is to monitor appropriate risk management policies, procedures and systems by collaborating with risk management;

DirectionRisks: it is responsible for setting up risk indicators and controlling each activity of the bank that may trigger a risk. It is therefore composed of a risk management department;

ManagementRisks : activates the overall risk management process involving the identification, definition, measurement and monitoring of all risks;

FinancialDirection :it checks the availability of the necessary financial resources for the management of the risk and subsequently it can allocate its resources to the risk management;

AuditControl : reviews internal procedures, assesses risk management tools and measures reliability and integrity to generate systematic and objective assessment reports;

EntityROJIH: attached to the direction risks, it is mainly concerned by operational, juridical IT and human risks;

ActionPlan: it identifies actions to reduce the impact of risks.

4.5 Identification of performance indicators

The choice of indicators is the most delicate step as it relates to the objectives of decision-makers. These indicators should monitor the organization's evolution, facilitate decision-making and improve performance. The selected indicators should be simple and representative. This choice was based mainly on the bibliography and the specific problems of this project. As a result, seven organizational performance indicators were identified:

Team Mobilization : this indicator enables to count the number of people present on each committee of the risk management unit since each committee is composed of a limited number of persons;

Means: identifies the means available at the time of risk. Each committee has its own means to intervene in risk management;

Duration: is the duration of mobilization of the unit actors. It generally equals one working day except for the cases of risk with very high degree;

Quality of information: when managing a risk, the key element is the information that will make the right decision at the right time. If the information circulates with difficulty or not entirely, the decision-maker will not be able to make the right decision;

The actors' competence: it allows knowing if all the actors are qualified and competent in the performance of their tasks. It is measured by the actors training within the bank. This is a central element used to increase their performance;

The group Cohesion: it relies on actors' motivation, shared emotions and interpersonal relationships;

The actors Morale: it can be affected at risks and it will surely have repercussions on the performance of the bank since the psychic state of the actors has a strong influence on the management of risks.

5 Elaboration of a prototype

We propose a system that evaluates the risk unit by taking into account the performance indicators chosen beforehand. It is equipped with an SMA that allows simulating the different scenarios.

The developed prototype consists of three interconnected parts (Fig. 3):

The user interface that allows adding scenario details of a risk to analyze the results in the form of a dashboard;

Functional part that simulates the risk scenario. It is responsible for the collection and processing of information;

Last part which constitutes the database for storing the information necessary for the simulation.

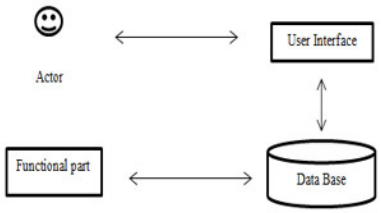


Figure. 3. Technical architecture

5.1 The user interface

The user interface is developed using the Java language under the Netbeans environment. It is connected to the database.After authentication through a login and a password, since it is the only one capable of simulating scenarios, the manager goes on to enter the data of the simulation.

5.2 The functional part

The role of this part is to detect the dysfunctions of the organization using the multi-agent systems according to a specific scenario. The system consists of a set of autonomous agents that each have separate threads and communicate with each other by sending messages. The MAS model of this system is represented by Fig. 4.

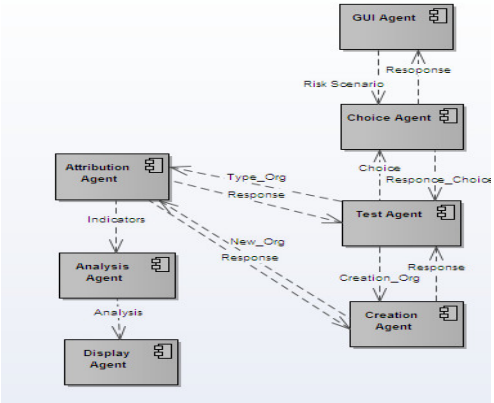


Figure .4. MAS structure of the proposed system

GUI Agent: which retrieves the specific information from the risk scenario via the user interface and seeks for a treatment from the choice agent;

Choice Agent: which allows the user to choose one unit organization among several according to the specificities and actors involved in risk management;

Test Agent: retrieves the user's opinion on whether he agrees on the type of organization or wants to add a new type;

Creation Agent: in case the user does not agree with the proposed organization, the agent proposes to him to define the type of the organization that suits him and is in charge of its creation;

Attribution Agent: which assigns to the actors of the chosen organization the value of the performance indicators which can be either 0 or 1 through a graphical interface;

Analysis Agent: Which is responsible for all the analyzes concerning the estimation of the duration of the simulation, the verification of the indicators and the study of the evolution of each indicator;

Display Agent: This displays the results of simulation.

5.3 The database

Data storage will be done using a database. We chose to work with SQL Lite because the recorded information is not large.

6 Conclusion

The aim of this paper was to propose a new approach for analyzing and evaluating the organizational performance of a risk management unit. The main idea is to simulate the process of risk management based on the systemic thinking and artificial intelligence. The result is a multi-agent system that integrates the performance indicators to simulate each type of risk within banks. The system proposed is being deployed and its uses will allow bank to improve their performance by analyzing the process of risk management and to change the internal structure.

REFERENCES

[1] P. A. Chiappori et M. O. Yanelle, «Le risque bancaire : un aperçu théorique,» Revue d'économie financière , pp. 97-111, 1996.

[2] O. De Bandt, P. Hartmann et J. L. Peydró, Systemic Risk in Banking, The Oxford Handbook of Banking, 2012.

[3] A. Mouhtadi, «Risque de liquidité : cas du Maroc,» International Journal of Innovation and Applied Studies, vol. 10 No, pp. 325-335, 2015.

[4] M. A. Malina et F. H. Selto, «Choice and change of measures in performance measurement models,» Management Accounting Research, vol. 15, pp. 441-469, 2004.

[5] O. Villarmois, «Le concept de performance et sa mesure : un état de l'art, Les Cahiers de la Recherche, Centre Lillois d'Analyse et de Recherche sur l'Evolution des,» 2001.

[6] M. Karsky, La dynamique des systèmes complexes ou la systémique de l'ingénieur, 2004.

[7] S. Sheard, «Complex Systems Science and its Effects on Systems Engineering,» chez European Systems Engineering Conference, 2006.

[8] F. Le Gallou, «Systémique: Théorie et applications,» chez Editions Tec Et Doc, 1993.

[9] G. Donnadieu and M. Karsky, La systémique, penser et agir dans la complexité, Paris: Editions de liaisons, 2002.

[10] J. M. Flaus, E. Berthelier et F. Giannoccaro, «Modélisation de systemes organisationnels pour l'analyse des défaillances : application au plan de sauvegarde communal,» chez Conférence Internationale de Modélisation et Simulation, Tunisie, 2010.

[11] . G.-M. Karagiannis, thèse: Méthodologie pour l'analyse de robustesse des plans de secours industriels, 2010.

[12] J. Flaus , E. Berthelier et F. Giannoccaro, «Modélisation de systemes organisationnels pour l'analyse des défaillances : application au plan de sauvegarde communal,» 2010.

[13] J. Lin, L. Chaoyu Lin et S. Huang, «Improving System Performance by Extending Inheritance Analysis and Design,» 2016.

[14] M. Bienvenu, «Instructor for online course ''Introduction to Artificial Intelligence'' Institution: Université de Provence, Marseille, France,» 2006-2008.

[15] F. Bellifemine, A. Poggi and G. Rimassa, «JADE – A FIPAcompliant agent framework,» 1999.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Feature-rich PoS Tagging through Taggers Combination : Experience in Arabic

Imad Zeroual and Abdelhak Lakhouaja
Computer Sciences Laboratory, Faculty of Sciences, Mohammed First University, Oujda, Morocco
mr.imadine@gmail.com, abdel.lakh@gmail.com

ABSTRACT

Since words can play different syntactic roles in different contexts, it is not trivial to assign the appropriate morphosyntactic category to each word according to the context. Part of Speech (PoS) tagging is the task which manage this issue. Several probabilistic methods have been adapted for PoS tagging such as Hidden Markov Models, Support Vector Machines, and Decision Tree. Based on these methods, language-independent PoS taggers have been developed such as TnT, SVMTool, and Treetagger. The main purpose of this work is to combine automatically the output of these standard PoS taggers and investigate several options for how to do this combination. The experiments are applied to one of the morphologically complex languages, Arabic. In this paper, we highlight the use of these taggers via various experiments. In fact, the evaluations involve several tests on both Classical and Modern Standard Arabic, trained/untrained and tagged/untagged data. Finally, a deeper investigation of Arabic PoS tagging through these language-independent taggers combination is performed.

Keywords: Part of speech, Tagging, Treetagger, SVMTool, TnT, Arabic.

1 Introduction

The Part of Speech tagging is the basis of well-known natural language processing (NLP) fields. It is a preliminary task that reflects directly the performance of any other subsequent text processing [1]. Further, pos tagging is a key input feature for other NLP tasks. For example, it is very useful for spell checking and correcting, named entity recognition, information retrieval, building dictionaries, phrase chunking, and text-to-speech synthesis systems [2].

Due to these convincing reasons, several methods have been proposed for automatic pos tagging. The most known are stochastic methods that used Hidden Markov Models (HMM), Support Vector Machines (svms), and Decision Tree (DT). Based on these methods, language-independent pos taggers have been developed. The finest freely available taggers with a considerable accuracy are tnt [3], svmtool [4], and Treetagger [5]. These taggers are data-driven which mean they learn from pre-annotated corpora and lexicon.

Arabic is a resource-poor language when it comes to finding freely available lexical resources and pre-tagged training corpora and so on. De facto, not a single tagged modern standard Arabic corpus was a freely or publicly available until the fall of 2011 [6]. Further, only a few works suggested a tagset for a

DOI: 10.14738/tmlai.54.2981
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2981

standard use such as [7], [8], and [9]. Thus, to evaluate different approaches and tools on a common ground becomes hard to realize.

Basically, the standard taggers are adaptable to any language if a lexicon and a tagged training data are available. Fortunately, a recent adaptation of Treetagger for Arabic [10] is available for public . The adaptation used a universal common tagset that covers 22 different languages including Arabic [11]. In this regard, we adopt the same language model used for the Treetagger to adapt the other two taggers (TnT and SVMTool).

To construct a combination system, two or more individual taggers should be involved. Usually, a combination of taggers obtains a higher accuracy than the application of just a single tagger [12]. The reason is that different taggers eventually produce different errors and these differences can be exploited to yield better results. Thus, when building combined taggers it is important to use taggers based on different methods [13].

The aim of the present paper is to shed light on the performance of three selected taggers via various experiments. Further, an evaluation of these taggers and their combinations is performed by measuring the accuracy rate at the word level. The corpora used for the experiments are; those which the taggers trained on, Al-Mus'haf [14] and NEMLAR [15]. Al-Mus'haf corpus covers the Quranic text where all the words are annotated using a semi-automatic method by applying "AlKhalil Morpho Sys" [16] and hand-corrected by linguistics afterwards. NEMLAR is an Arabic written corpus produced and tagged by RDI, Egypt for the Nemlar Consortium. The third data is untagged and extracted from the Arabic part of a multilingual corpus [17]. Therefore, the tagging results were validated manually and compared to their combination.

Finally, all possible options for how to do these dual combinations are performed and their results is discussed according to their individual performance. The obtained results prove that some of these combinations could have a negative effect on the system performance, but others yield a better result. However, the best accuracy is achieved by extending the dual combination to a trilogy.

The remainder of the paper is organized as follows: in the second section, background information is given about Arabic PoS tagging, mentioning several relevant works. We briefly describe the three standard taggers and the universal tagset implemented in the third section. In the fourth section, various experiments are presented and we discuss the obtained results. Finally, we conclude this paper in section five.

2 Arabic PoS tagging

The PoS tagging is selecting which is necessary to resolve ambiguity during text processing to estimate the exact morphosyntactic tags to fit the actual input context.

In order to begin with PoS tagging, There are certain requirements [18]:

To select the approaches that will be used for the automatic tagging process;

To prepare the linguistic resources required for training the tagger, and optional lexicon containing all possible tags for a particular word form;

To define the tagset, i.e., basic morphosyntactic tags attached to each word.

Different methods are designed for PoS tagging:

Statistical methods: they derive the probabilities from a large pre-tagged training data. The majority of these methods are based on HMM. For example, Banko and Moore [19] presented an HMM tagger that exploits context on both sides of a word to be tagged. It has been evaluated in both the unsupervised and supervised cases. It achieves an accuracy of about 96%. Another tagger has been proposed, tested, and it achieves a performance of 97% [20]. The last examples is a recent published study by [21]. Other statistical taggers are based on SVM such as [22] and Yamcha [23] that achieves 97.6% accuracy. Finally, a maximum entropy approach, that enrich the information sources used for tagging, has been adopted by the Stanford PoS tagger [24]. Its result accuracy on the Penn Treebank [25] achieves 96.86% overall, and 86.91% on previously unseen words.

Rule-based methods: they are based on rules defined by linguistics. These methods involve morphological analysis and lexicons research. For example, Brill's PoS tagger, a rule-based tagger for Arabic which uses a machine learning approach based on Brown corpus [26]. A similar work has been developed by [27] using transformation-based learning method, which is an error-driven approach to induce the retagging rules from a training Arabic Treebank corpus and the morphological analyser BAMA [28]. The achieved accuracy is 96.9%. Finally, an Arabic PoS tagger has been developed based on sentence structure, i.e., the relation between the untagged words and their adjacent [29].

Neural network models: These methods use learning models inspired from the artificial intelligence field based on understanding the operation of biological neural networks in brains. Typically, they use highly interconnected simple processing nodes [30]. To our knowledge, no implementation of neural network models has been done for Arabic yet. As Carneiro claimed [31], the developer of the mWANN-tagger (multilingual Weightless Artificial Neural Network tagger), the ability to successfully tag languages that possess non-concatenative morphology such as Arabic is left for a future investigation.

Hybrid Systems: guessing PoS tags for unknown word is certainly the main problem in probabilistic methods. This problem becomes more serious in resource-poor languages that have rich vocabulary and complex morphology such as Arabic [32]. In order to handle this problem, several taggers combine different statistical model with rule-based method like [33], [34], and [35]. Their accuracies achieved are respectively 97.4%, 94%, and 92.86%. It should be noted that the three taggers differ in terms of the selected tagset and the resources used in both training and evaluation. A different hybrid system has been implemented by Tlili-Guiassa [36] to tag an Arabic text based on memory-based learning and rules-based method. The modus operandi of this system is to apply rules (analyzing the affixes and the patterns of a word) to determine the appropriate tag of each word in the current context, then, refer to memory-based learning as a machine learning method that can handle the exceptions of these rules.

3 Standard taggers

Standard taggers have been successfully applied to several morphologically complex languages such Arabic, which have been yielded state-of-the-art results. In the following, we introduce those most relevant ones:

- TnT tagger: TnT (short form of Trigrams'n'Tags) is a PoS tagger developed by Brants [3]. TnT tagger uses the Viterbi algorithm [37] for second-order Markov models where the transition probability depends on two preceding tags. From a training data, the probabilities of the model are estimated using maximum likelihood estimation. New

assignments of tags to a word are determined by optimizing the product of lexical probabilities and contextual probabilities. The states of the model represent tags while the outputs represent the words.

- Treetagger: it uses an unknown word PoS guesser similar to that of the TnT tagger. However, it is developed to avoid problems that HMM face in transition probabilities. Thus, Treetagger estimates the transition probabilities based on a decision tree; which mean that the probability of a given trigram is determined by following the corresponding path through the tree until a leaf is reached [5]. The Treetagger is probably the widely language-independent PoS tagger used, it has been officially and successfully used to tag more than 30 different languages.
- SVMTool: Giménez and Marquez [4] proposed a standard PoS tagger based on SVMs. the tagger implement five different models for training (0 (default),1, 2, 3, and 4) with a tagging direction that can be either “left-to-right”, “right-to-left”, or a combination of both. Concerning the models training, they are based on the SVMLight implementation of the Vapnik's SVM [38], [39] by Joachims [40].

All the three taggers come with two programs. The first one is for the training task that requires a tagged training corpus and auxiliary lexicon to generate the parameter file. The second program is the tagger itself. Up to date, there is no comparative study that evaluates the three taggers for Arabic on a common ground. However, the taggers have been individually implemented and evaluated for English under the same conditions. They were trained on two million words of the Wall Street Journal section of the Penn Treebank corpus [41]. The obtained accuracy rates are 96.06%, 96.36%, and 97.16% respectively for the TnT, Treetagger, and SVMTool [4], [42]. This shows that the SVMTool outperforms the other taggers.

3.1 The tagset

A tag is a string used as a label to represent information about morphosyntactical features (case, gender, etc.) of word forms. A tagset is a set of these tags.

Generally, every proposed PoS tagger is looking for his suitable objective. Therefore, it is hard to assume that the tagsets proposed up to date are for standard use; especially for a language with a highly inflectional grammar, Arabic. Since we don't want to reinvent the wheel, the recent adaptation of Treetagger, that use a tagset based on two studies [11], [43] is thus adopted in our experiments. Table 1 below presents the basic tags of the adopted universal tagset.

Table 1. The basic tags of the universal tagset.

Tags	Tag Symbols	Tag in Arabic	Examples
1. Verbs (all tenses and modes) 2. Nouns 3. Proper nouns 4. Pronouns 5. Adjectives 6. Adverbs 7. Utilities words (Particles, Adpositions…) 8. Disconnected letters (Quranic Initials) 9. Speech-specific sounds 10. Other: foreign words, typos, abbreviations… 11. Punctuation marks	VERB NOUN PN PRON ADJ ADV PRT DISL Uh X SENT	فعل اسم اسم علم ضمير صفة ظرف أداة حروف مقطعة حرف صوت أخرى علامة ترقيم	"كَتَبَ" (kataba "to Write") "مَدْرَسَة" (madrasap "School") "مُحَمَّد" (muHam~ad "Mohamed") "هِيَ" (hiya "She") "جَمِيل" (jamyl "Beautiful") "بَعْدَ، فَوْقَ" (baEda, fawoqa "After, Above") "إلى، ذلك، الذي" (<ilY, *lk, Al*y "To, That, who") "الم، طه، كهيعص" (Alm, Th, khyES) "آه، هيهات" (|h, hayhAt) "أوبك، مانشستر" (>wbk, mAn$str "OPEC, Manchester")

4 Experiments and discussion

In this section, we highlight the performance of the selected taggers via various experiments on text from both Modern Standard Arabic (MSA) and Classical Arabic (CA). Further, we evaluate the performance of the taggers on trained/untrained and tagged/untagged data. In addition, all possible combinations are investigated, presenting the best combination. Finally, the achieved results are discussed.

4.1 Experiences on tagged/trained corpora

In this section, we highlight the performance of the three taggers (TnT, Treetagger, and SVMTool) and explore the commons results and different errors. At the first stage, these taggers are evaluated individually using both tagged and trained corpora. The taggers are trained and tested on data from the NEMLAR (500,000 words) and Al-Mus'haf (78,121 words) corpora. 90% of the corpora are used for training and the rest 10% for testing. Table 2 exhibits the obtained accuracies from all the taggers

Table 2. Accuracy results

* WORDS ARE NOT INCLUDED IN TRAINING DATA

Corpora	*Unrecognized words**	*TnT*	*Treetagger*	*SVMTool*
Al-Mus'haf	942	93.97%	**94.70%**	93.42%
NEMLAR	6,276	94.74%	**95.12%**	94.88%

As seen in Table 2, Treetagger performs better than the other taggers when they are applied on both corpora. Whereas, the achieved accuracy by TnT is slightly better than the one achieved by SVMTool when it is applied on Al-Mus'haf corpus; and vice versa when they are applied on NEMLAR corpus.

To indicate the motivation for taggers combination, deeper investigation is required. Therefore, we checked the outputs of the three taggers to explore the commons results and different errors, and eventually to exploit these observations in further tasks. Table 3 shows detailed information about taggers outputs.

Table 3. Detailed information about taggers outputs

Corpora	***Common***		***Non-common***			
Taggers	*All*		*TnT*	*Treetagger*	*SVMTool*	*All*
Al-Mus'haf	*93.72%*		*6.28%*			
	Correct	*Incorrect*	*Correct*	*Correct*	*Correct*	*Incorrect*
	92.94%	0.78%	1.03%	1.76%	0.48%	3.01%
NEMLAR	*94,61%*		*5.39%*			
	Correct	*Incorrect*	*Correct*	*Correct*	*Correct*	*Incorrect*
	93.85%	0.76%	0.89%	1.27%	1.03%	2.20%

Several hints are observed in Table 3 above:

- The common outputs are not certainly correct; yet, the rate of incorrect ones remains very low (0.76%-0.78%).
- None of the common and correct outputs of the three taggers reach the accuracy rate of the three taggers individually.
- The non-common outputs are not certainly incorrect. In fact, more than half of them are correct.
- Based on these observations, we deduce that depending only on the common outputs is not effective, because it does not reach the performance level of each tagger individually. Also, we cannot abandon the non-common outputs, where there is an interesting percentage of correct results.

4.2 Combination algorithm

Through the previous investigation, it is possible to define an appropriate combination algorithm. However, the purpose of this work is not to propose a better combination algorithm. Instead, we would like to demonstrate that a combination system does effectively improve tagging accuracy. Here, we describe the algorithm implemented for the combination process. This combination algorithm determines the most appropriate tags in three steps:

- Tagging the input text with all taggers;
- Selecting for each token the most voted tag from the majority taggers (in these experiments, at least two taggers);
- If the given tags from all taggers are unlike. Then, the selected tag is the one proposed by the most accurate tagger (in these experiments, is Treetagger).

The evaluation of the algorithm is divided into two phases. In the first one, only two taggers are used as a combination. Consequently, we left with three possible combinations; while in the second phase, the

three taggers are used as a combination. Table 4 shows the achieved accuracies of all combinations in these the two phases.

Table 4. Combinations accuracies

Combinations	*TnT & Treetagger*	*Treetagger & SVMTool*	*TnT & SVMTool*	*All taggers*
Al-Mus'haf	95.73% (+)	93.54% (-)	93.82% (-)	**95.79% (*)**
NEMLAR	95.23% (+)	95.00% (-)	94.93% (+)	**96.45% (*)**

By combining the outputs of two or three taggers using the proposed algorithm, the obtained results are as follows:

- (-): these combinations achieve an accuracy rate lower than the most accurate tagger involved in the combination.
- (+): these combinations achieve an accuracy rate higher than the most accurate tagger involved in the combination.
- (*): the best achieved results in all combinations, i.e., those involve all the three taggers.

4.3 Evaluation on untagged/untrained corpus

After testing and validation of the combination algorithm on available pre-tagged and trained corpora, it remains to evaluate this algorithm on a new untagged/untrained data which is the main objective of this work. For that reason, we have selected the data from a rich resource in term of variety of domains and topics. The data are extracted from the Arabic part of a new proposed multilingual corpus constructed based on the available subtitles of TEDx talks. The size of these data is 500,000 words.

Before applying the combination algorithm, it is required to determine the most accurate tagger among the three. Based on the idea that the tagger which outperforms the others in the non-common outputs, eventually, it will be the one that has the higher accuracy in the overall corpus. Hence, the first stage of this evaluation is to annotate the corpus with the three taggers to separate the common and non-common outputs. Finally, we verify manually and validate the achieved accuracies in two experimental samples: (1) all non-common outputs; (2) 10% of random common outputs. Table 5 presents the obtained results of this task.

Table 5. Accuracy analysis on experimental samples

	Common		**Non-common**			
Taggers	**All**		**TnT**	**Treetagger**	**SVMTool**	**All**
Percentages	86.98%		13.02%			
Experimental samples	10.00%		13.02%			
Correctness	**Correct**	**Incorrect**	**Correct**	**Correct**	**Correct**	**Incorrect**
Accuracy	84.85%	2.13%	**4.03%**	4.02%	3.54%	1.43%

The observations noted in the previous evaluations (Table 3) remain the same as they are in the experimental samples presented in Table 5. Thereby, the next step is to apply the combination algorithm and compare it to the performance of each tagger. In this task, only both experimental samples (23.02%)

are used instead of the overall corpus. Table 6 exhibits the obtained accuracy rate by all the taggers and the combination algorithm.

Table 5. Taggers accuracies on TED corpus

Taggers	*TnT*	*Treetagger*	*SVMTool*	*Combination*
Accuracy	88.88%	88.87%	88.39%	**90.63%**

As seen in Table 6, the achieved accuracies of the three taggers are approximately similar with relative progress of the TnT tagger. However, the combination algorithm outperforms the three taggers individually.

4.4 Discussion

Several experiments on various resources are performed in term of text form, trained/tagged and untrained/untagged. Based on a deeper investigation of these experiments, a combination algorithm is developed. Several evaluations and validations are done to demonstrate that the combination system does effectively improve tagging accuracy considering the number of taggers involved and their performance. To sum up the most important results obtained in this investigation, we state the following points:

- As seen in the evaluation experiments, the proposed combination system performs better than the other taggers when they applied individually on all three corpora.
- Usually, the PoS tagging is done by an automatic process and manually corrected afterwards. To minimize the hand-correction, the combination algorithm can be used to improve the accuracy rate; yet, to point the candidate mis-tagged words by indicating the unlike tags.
- By combining only two taggers, an accuracy rate reduction could be achieved. In our case, the rate was lower than the most accurate tagger involved in the combination algorithm.
- Improving the performance of the current combination algorithm is at hand. For instance, the improvement still possible if the number of involved taggers is augmented or different combination algorithms are adopted.

In addition, all obtained results shows that the accuracy of a common output is always lower than that achieved by the taggers separately. The reason is that the taggers produce different errors and these differences are exploited in the combination to yield better results. Therefore, we suggest combining taggers based on different methods while building a combined system.

5 Conclusion and perspectives

In this paper, we demonstrate the feature-rich functionality of PoS tagging through taggers combination. Arabic language was the case of this investigation. The combination algorithm achieves a state-of-the-art overall accuracy in Arabic PoS tagging and outperforms other taggers.

Here, we highlight pertinent tagging methods, primarily those implemented for Arabic language, mentioning relevant works that have been published in this field. Further, three standard taggers are introduced with a brief description; then, their performance is evaluated via various experiments by measuring the accuracy rate at the word level.

As it was observed in the experiments conducted, the proposed combination algorithm involves three language-independent PoS taggers. The performance is better in comparison with the other taggers separately. The proposed system increases the accuracy rate of the most accurate tagger by 1.09%, 1.33% and 1.75% respectively for the three different corpora Al-Mus'haf, NEMLAR and the Arabic part of TED corpus. The modus operandi of our PoS combination algorithm is in accordance with the observations concluded during the various experiments that we have made.

Finally, this work is another step to improve tagging accuracy for Arabic and to minimize the hand-correction. Yet, it is still possible to move performance levels up. Later, we look forward to combine a PoS taggers dedicated to the Arabic language (e.g., [34], [44], and [45]) with an application to fine-grained PoS tagging. Further, we plan to investigate other combining approaches and their application to PoS tagging.

REFERENCES

[1] M. Albared, N. Omar, and M. J. Ab Aziz, "Developing a competitive HMM arabic POS tagger using small training corpora," in Intelligent Information and Database Systems, Springer, 2011, pp. 288–296.

[2] R. A. Abumalloh, H. M. Al-Sarhan, O. Ibrahim, and W. Abu-Ulbeh, "Arabic Part-of-Speech Tagging," J. Soft Comput. Decis. Support Syst., vol. 3, no. 2, pp. 45–52, 2016.

[3] T. Brants, "TnT: a statistical part-of-speech tagger," in Proceedings of the sixth conference on Applied natural language processing, 2000, pp. 224–231.

[4] J. Giménez and L. Marquez, "SVMTool: A general POS tagger generator based on Support Vector Machines," 2004.

[5] H. Schmid, "Treetagger| a language independent part-of-speech tagger," Inst. Für Maschinelle Sprachverarbeitung Univ. Stuttg., vol. 43, p. 28, 1995.

[6] H. S. Rabiee, "Adapting Standard Open-Source Resources To Tagging A Morphologically Rich Language: A Case Study With Arabic.," in RANLP Student Research Workshop, 2011, pp. 127–132.

[7] S. Alqrainy, "A morphological-syntactical analysis approach for Arabic textual tagging," De Montfort University, 2008.

[8] M. Sawalha, "Arabic Morphological Features Tag set." University of Leeds, 2009.

[9] I. Zeroual, A. Lakhouaja, and R. Belahbib, "Towards a standard part of speech tagset for the Arabic language," J. King Saud Univ. - Comput. Inf. Sci., 2017.

[10] Z. Imad and L. Abdelhak, "Adapting a decision tree based tagger for Arabic," in 2016 International Conference on Information Technology for Organizations Development (IT4OD), 2016, pp. 1–6.

[11] S. Petrov, D. Das, and R. McDonald, "A universal part-of-speech tagset," ArXiv Prepr. ArXiv11042086, 2011.

[12] N. P. P. Khin and T. N. Aung, "Analyzing Tagging Accuracy of Part-of-Speech Taggers," in International Conference on Genetic and Evolutionary Computing, 2015, pp. 347–354.

[13] V. Henrich, T. Reuter, and H. Loftsson, "CombiTagger: A System for Developing Combined Taggers.," in FLAIRS Conference, 2009.

[14] I. Zeroual and A. Lakhouaja, "A new Quranic Corpus rich in morphosyntactical information," Int. J. Speech Technol., pp. 1–8, Feb. 2016.

[15] M. Yaseen et al., "Building annotated written and spoken Arabic LR's in NEMLAR project," in Proceedings of LREC, 2006, pp. 533–538.

[16] M. Boudchiche, A. Mazroui, M. Ould Abdallahi Ould Bebah, A. Lakhouaja, and A. Boudlal, "AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer," J. King Saud Univ. - Comput. Inf. Sci., 2016.

[17] I. Zeroual and A. Lakhouaja, "Towards a Multilingual Aligned Parallel Corpus," presented at the International Conference of High Innovation in Computer Science, Kenitra, Morocco, 2016, pp. 1–4.

[18] M. Utvić, "Annotating the corpus of contemporary Serbian," in Proceedings of the INFOtheca '12 Conference, 2011.

[19] M. Banko and R. C. Moore, "Part of speech tagging in context," in Proceedings of the 20th international conference on Computational Linguistics, 2004, p. 556.

[20] F. Al Shamsi and A. Guessoum, "A hidden Markov model-based POS tagger for Arabic," in Proceeding of the 8th International Conference on the Statistical Analysis of Textual Data, France, 2006, pp. 31–42.

[21] A. Kadim and A. Lazrek, "Bidirectional HMM-based Arabic POS tagging," Int. J. Speech Technol., vol. 19, no. 2, pp. 303–312, 2016.

[22] M. Diab, K. Hacioglu, and D. Jurafsky, "Automatic tagging of Arabic text: From raw text to base phrase chunks," in Proceedings of HLT-NAACL 2004: Short Papers, 2004, pp. 149–152.

[23] N. Habash and O. Rambow, "Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop," in Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005, pp. 573–580.

[24] K. Toutanova and C. D. Manning, "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger," in Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13, 2000, pp. 63–70.

[25] M. Maamouri and A. Bies, "The Penn Arabic Treebank." In Farghaly, A., Ed., Arabic Computational Linguistics. CSLI Publications, Stanford, CA., 2010.

[26] A. Freeman, "Brill's {POS} tagger and a Morphology parser for {Arabic}," 2001.

[27] S. AlGahtani, W. Black, and J. McNaught, "Arabic part-of-speech tagging using transformation-based learning," in Proceedings of the Second International Conference on Arabic Language Resources and Tools, Cairo, Egypt, 2009.

[28] T. Buckwalter, "Buckwalter Arabic morphological analyzer (BAMA) version 2.0. linguistic data consortium (LDC) catalogue number LDC2004L02," ISBN1-58563-324-0, 2004.

[29] Y. El Hadj, I. Al-Sughayeir, and A. Al-Ansari, "Arabic part-of-speech tagging using the sentence structure," in Proceedings of the Second International Conference on Arabic Language Resources and Tools, Cairo, Egypt, 2009.

[30] D. R. Wilson, "Advances in instance-based learning algorithms," Citeseer, 1997.

[31] H. C. Carneiro, F. M. França, and P. M. Lima, "Multilingual part-of-speech tagging with weightless neural networks," Neural Netw., vol. 66, pp. 11–21, 2015.

[32] M. Albared, T. Al-Moslmi, N. Omar, A. Al-Shabi, and F. M. Ba-Alwi, "Probabilistic Arabic part of speech tagger with unknown words handling," J. Theor. Appl. Inf. Technol., vol. 90, no. 2, p. 236, 2016.

[33] M. Hadni, S. A. Ouatik, A. Lachkar, and M. Meknassi, "Hybrid Part-Of-Speech Tagger for Non-Vocalized Arabic Text," Int. J. Nat. Lang. Comput., vol. 2, no. 6, pp. 1–15, 2013.

[34] N. Ababou and A. Mazroui, "A hybrid Arabic POS tagging for simple and compound morphosyntactic tags," Int. J. Speech Technol., vol. 19, no. 2, pp. 289–302, 2016.

[35] A. H. Aliwy, "Arabic Morphosyntactic Raw Text Part of Speech Tagging System," Repozytorium Uniwersytetu Warszawskiego, 2013.

[36] Y. Tlili-Guiassa, "Hybrid method for tagging Arabic text," J. Comput. Sci., vol. 2, no. 3, pp. 245–248, 2006.

[37] D. L. Neuhoff, "The Viterbi algorithm as an aid in text recognition (Corresp.)," Inf. Theory IEEE Trans. On, vol. 21, no. 2, pp. 222–226, 1975.

[38] V. Vapnik, The nature of statistical learning theory. Springer Science & Business Media, 2013.

[39] V. N. Vladimir and V. Vapnik, The nature of statistical learning theory. Springer Heidelberg, 1995.

[40] T. Joachims, "Svmlight: Support vector machine," SVM-Light Support Vector Mach. Httpsvmlight Joachims Org Univ. Dortm., vol. 19, no. 4, 1999.

[41] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini, "Building a large annotated corpus of English: The Penn Treebank," Comput. Linguist., vol. 19, no. 2, pp. 313–330, 1993.

[42] H. Schmid, "Probabilistic part-ofispeech tagging using decision trees," in New methods in language processing, 2013, p. 154.

[43] O. Rambow et al., "Parallel syntactic annotation of multiple languages," in Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC2006). Genoa, Italy, 2006.

[44] A. Pasha et al., "Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic," in Proceedings of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.

[45] K. Darwish, A. Abdelali, and H. Mubarak, "Using Stem-Templates to Improve Arabic POS and Gender/Number Tagging.," in LREC, 2014, pp. 2926–2931.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

The Rate of Local Integration in the Supply Chain of the Renault Maroc Company

Rajaa Raissouni ,M'hamed Hamiche
Centre des Etudes Doctorales en Economie et Gestion et Développement Durable
de l'Universite Abdelmalek Esaadi, Tangier, Morocco
rajaaraissouni@gmail.com;Hamiche2020@gmail.com

1 Introduction

The automobile industry has become Morocco's leading export market and one of the most successful industrial pillars in the Moroccan industry. First with the establishment of Renault Tanger in 2008, which is the biggest foreign investment in Moroccan history, furthermore with the anticipated arrival of Peugeot-Citroën (PSA) in Kenitra in 2019; therefore, Morocco remains an attractive country for major car makers due to its political stability and favorable geographical position compared to its neighbors in the region.

The development of a solid industrial fabric is the result of further integration of the automotive value chain in Morocco. Indeed, the year 2016 testifies to the signing of supplier agreements called Ecosystem, mobilizing investments of the order of 10 billion dirhams.

These agreements were signed in the presence of His Majesty King Mohamed IV and several ministers as well as the Director of Operations, of the Renault group Middle East and India branch. This project involves developing a global supply platform that generates a figure of 20 billion dirham per year. It will contribute to the creation of 50,000 new permanent jobs with the objective of doubling this number by 2020 to more than 160000. In short, the ecosystem signed in April 2016 includes three types of agreements:

- A framework agreement on the establishment of the Renault ecosystem.
- An industrial property agreement.
- A convention on executive training and vocational training in the automotive sector.

Morocco commits to give subsidies to investments and tax advantages to suppliers who intend to settle within the country. Renault took advantage of this ecosystem agreement by increasing its local integration rate and consequently reducing and optimizing its logistics and purchasing costs.

Generally, the purchasing costs of multinationals such as Renault represent a significant percentage of the cost of their products. To this end, they still hope to optimize them by several levers, local sourcing is a strategic tool to achieve this objective through its direct contribution to the purchasing performance and consequently the company's performance.

This leads us to question the levers and the challenges that this local sourcing represents for the major international structures. In other words, the mean question is: does the increase in the rate of local integration always reduce purchasing costs for multinationals?

DOI: 10.14738/tmlai.54.2982
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2982

Keywords: Supply chain; Local integration rate ; Automobile industry; Sourcing procedures; sourcing practice.

2 Sourcing strategies

Sourcing is the research, location, selection and evaluation of a supplier in order to meet an identified need. However, in a context of globalization, multinationals manage sourcing around the world by managing a global supply chain for their production sites scattered across the globe. To do this, multinationals seek to reduce the logistical costs of their factories by looking for locally installed suppliers who can supply and deliver according to their specific standards.

2.1 Sourcing strategies according to the supplier panel:

2.1.1 The different sourcing strategies

Sole sourcing:

It is a single source of supply for one type of product. It can be the result of sourcing from a supplier whose production capacities are rare in the market; But it would be wise to look at the supplier market to compare the supply of other suppliers with the one already contracted.

The use of a single supplier presents a strategic risk insofar as an incident can lead to a break in the supply chain, the consequence of which can be damaging for the enterprise. The link between sole sourcing and supplier relations gives rise to a less formal and longer-term contact. But having a sole sourcing strategy to increase quality is not always true[1] .

Dual sourcing:

It involves using two suppliers to produce the same component or buy the same raw material. In general the business is shared for 70% with the supplier Y and 30% with the supplier X. It is in some way a response to the risk of the single source of supply since the supplier acts as a backup in case the other would be failing. Many companies consider dual sourcing as important because it allows the company to have some flexibility as well as avoid being frozen by long-term contracts with a single supplier.

Multi sourcing:

Multi sourcing is when three or more suppliers are used by the company for the same product. This corresponds to a strategy of diversifying sources of supply when, for example, a parent company orders from a number of suppliers close to its factories and markets a similar product to supply the local demand.

[1]Falguni Sen et William J. Latzko, Single Source Supplier Strategies—An Exploratory Study of the Impact of Values, Attitudes and Practices in Large U.S. Manufacturing Firms. Academy of management Proceedings, 1987, p.301-305

2.1.2 Back-up strategies:

Cross sourcing:

Cross sourcing is a sourcing strategy in which a company uses a supplier in the production of a good or service for a production domain of the company and another supplier producing the same good or service for another area of production line.

The idea is to have a back up supplier and develop a competition between the two suppliers for future contracts to be awarded. Competition between suppliers is a means of encouraging emulation but should not be an instrument of pressure of the client company on its supplier.

Suppliers of fluctuating demand:

Some industries experience a peak demand during a season of the year. They use a supplier that produces continuously during the year. During the strong period, one or more other suppliers will support the peak demand. The degree of strategic relationship with the temporary supplier is not necessarily very high. Sometimes it serves as a "spare wheel", thing that hampers to carry out a strategic purchasing policy with a reduction in prices and the appearance of innovation.

2.2 The Geo-Location of the Company And The Location Of Resources:

2.2.1 Local sourcing:

Local sourcing involves working with regional or national suppliers, defined according to the location of a business activity as a factory. Local sourcing has the advantage of reducing the risks associated with logistics. Another advantage is to benefit from a product delivered with speed, when the freshness of the products is critical or when the demand evolves quickly.

For a company that does not internationalize its sourcing, local sourcing provides, at a lower cost, a developed network of partner suppliers to cover the demand in either local or international markets. However, if these sourcing concerns manufactured product in a declining cycle, low-end, designed by unskilled employees, the local sourcing in certain regions is not advantageous because costs become the first criterion.

2.2.2 International sourcing:

International sourcing is possible through the development of transport and telecommunications which makes it more accessible. This international sourcing is used to benefit from a comparative advantage like that of Silicon Valley where innovations are strong because emulation is favored. Thus appears a geography of sourcing where the firms look for quality in some leading countries like France, Japan, Germany and the United Kingdom[2].

[2] AberdeenGroup, Low-Cost Country Sourcing : Success Strategies Maximizing and Sustaining the Next Big Supply Savings Opportunity, Juin 2005

International sourcing has developed with the emergence of competitive players outside the borders. With the collapse of the Soviet Union, China's openness, WTO agreements, the development of NAFTA, raw materials, plastics, electronics are bought more frequently from international sources.

This sourcing is adapted when, for manufactured products, the time taken for ship delivery is not paramount. Otherwise the airway remains a possibility to transport in the emergency but its use is very expensive. International sourcing requires taking into account[3] the total costs of delivery (transport, customs clearance, banking, insurance, quality inspection); Trade regulations, both restrictions and incentives; Time to Market (flexibility); Value-added services (a service which is offered by a supplier and which makes it possible to dispense with one more intermediary); Internet communication tools. The barriers are understanding of international procurement procedures (import licenses, certificates of origin), extended delays, cultural differences, currency risks. Finally, having international suppliers has the effect of increasing the inventory level resulting from a policy of risk reduction[4] (downtime, terrorism, change, policy) that leads to the accumulation of security stocks.

2.3 The sourcing procedure

2.3.1 Sourcing steps

Expenditure Analysis, Spend analysis:

Strategic sourcing begins with spend analysis up to the procurement automation process[5] . Spend analysis refers to cost management. It is a way of analyzing cost structures that can be carried out by product or by product groups. The products purchased and the procedures leading to these purchases are analyzed in great detail.

The proactive specifications:

Before a sourcing project, the company establishes a general project plan (mapping) in which it formulates its needs, proposes modes of action in the supplier market, makes internal recommendations, identifies the cost sources (cost drivers). It is the stage of constitution of the proactive specifications. This is the preparation stage prior to the decision to attack (as in marketing) the dense supplier market.

Even before the detailed specifications have been drawn up, it must precede a strategic decision by the company's top management. The intention is specified, communicated and comes into application by the constitution of a sourcing team. The latter will attempt to determine the buying segments where the proactive approach to the supplier market can be carried out. Once defined, the team draws up the complete specifications by consulting the other internal clients in order to guarantee the general approval in a strategy of global strategy known to all.

Data collection and processing:

Any sourcing project, whatever its strategic and geographic orientation, requires the company to collect and analyze all kinds of data in order to prepare the sourcing process and anticipate the risks. This may

[3] Sara Ireton, Global Sourcing Checklist. Industry Week, janvier 2007, p. 40

[4] Mike Jones, Domestic costs of foreign sourcing. Material Handling Management, février 2008, p. 40-41

[5] Sean Delaney, Delivering Value in both good and bad times, E-sourcing Forum, 14 juillet 2008

be data from previous sourcing operations: Results, quality of products purchased, and respect of deadlines by the supplier. This may include data on the economic environment such as inflation rates, world commodity prices, exchange rates and the transport network. Also the buyer is required to learn about the legislative environment. The latter can be very changeable and contain legal gaps in some countries; It is necessary for the firm to guard against possible risks due to non-protection by the law, such as intellectual property.

Generally it is necessary to collect the strategic information that will help to establish the sourcing decision, indeed it is necessary to look for 3 important information:

- Geo-location of supplier: this information will allow defining the logistic flows and the daily rates of delivery.
- Quality of supplier and price: An ISO certification of supplier is very important information that proves the quality of the products.
- Production capacity: knowing the supplier's production capacity will make it possible to know its competitiveness and its flexibility towards an unpredictable increase in demand, which generates large investments.

The offers received and selection suppliers:

Supplier selection is a critical phase in the sourcing process, as it has a significant impact on the company's performance.

Indeed, it is necessary to develop a network of reliable and competitive suppliers so that the company can raise and survive in a competitive environment. However, in order for the company to be able to provide a quality product at a reasonable cost, it must virtually monitor the performance of its suppliers.

However, there are two major problems:

- The determination of the number of suppliers and the mode of relationship with them: this choice is largely influenced by the strategic activity area of the company as well as its management mode,
- The selection of the best suppliers: this problem arises after the determination of the number of suppliers selected, however we will explain in detail in the next section the different techniques designed by researchers in the determination of the best suppliers and the fundamental criteria in the decision of making supplier appointment .

2.3.2 Supplier selection items:

Selection criteria:

Based on a survey of 274 Canadian and American firms that are members of the National Association of Purchasing Managers (NAPM) and in order to theorize supplier selection criteria, Dickson (1966) was able to identify 23 criteria used by companies in the 1960s to select their suppliers

The study showed that supplier selection is a multi-criteria decision that often involves simultaneous consideration of several criteria such as price, delivery time and quality, and that it is extremely difficult to find a supplier that excels everywhere. For example, the supplier offering the lowest price may not have the best performance in terms of delivery time or product quality.

Indeed, the criteria are classified by degree of importance, the table below, summarizes the different criteria and their classification:

Table 1: Dickson's classification of supplier selection criteria

Supplier Selection Criteria	Dickson rankings 1966
Price	1
dilevry	2
Quality	3
production Capacity	4
geographical Location	5
Technical capacity	6
Management and organization	7
Positioning in the Industry	8
Financial situation	9
Past Performance	9
Repair Service	9
Attitude	10
Packaging capacity	11
Operations Control	11
Training and support	12
Process Compliance	12
Social relationship	12
Communication system	12
Reciprocal relationship	12
Impression	12
Passion to do business	13
Purchase volumes	13
Warranty Policy	14

According to Vonderembse (1995), their study indicates that the process of supplier selection is multi-criteria and that the most important criteria are in this order: quality, product performance, reliability of delivery, product availability, cost, delay, technical capacity Supplier, after-sales service, financial situation and lastly the geographical location of the supplier.

The study also indicates that performance and product quality are the two most important criteria in selecting suppliers for companies, and recommends that firms reduce the number of suppliers and develop strategic partnerships with suppliers.

Types of relationships between the company and suppliers:

Masella and Rangoon (2000) grouped these relationships into two phases. Each variable depends on the horizon of the relationship and the degree of integration between the company and the supplier:

The horizon of the relationship:

We distinguish between a short-term relationship, the occasional (or cyclical) outsourcing and a long-term relationship, such as permanent (or structural) subcontracting. The duration of the relationship depends on several factors, the most important being:

- The level of investment in specific goods such as infrastructure, equipment, information systems, etc. These are the elements that the company must create to make the relationship operational and not to be used in other buying relationships..
- The cost of changing suppliers: It includes the cost of finding new suppliers in the case of low performance or the existence of a significant risk.

The degree of integration between the company and the supplier in logistics and strategic terms:

Logistics integration involves performance arrangements such as quality, service and time. However Strategic integration refers to arrangements that involve the supplier's know-how to develop new products and technologies.

Therefore, the type of relationship between an OD and a provider can be one of the following four scenarios A, B, C and D:

Table 2: Scenario of degree of enterprise-provider integration

	Short term	Long term
Logistics integration	A	B
Strategic integration	C	D

To determine the selection criteria for each of these scenarios, the supplier is modeled as a dynamic system, presented by three types of variables:

- Output variables: They correspond to the performance of the supplier and are of two classes: manufacturing performance such as cost, quality, delay, flexibility and service, and technological performance such as innovation of products purchased.
- Input variables: These include control variables used by the supplier to complete a certain performance and environmental variables such as market developments, competitor actions which are not completely under the control of the supplier.
- The state variables: They are responsible with the input variables of the dynamics of the output variables of the system which is the supplier. To define state variables, resource-based theory is used.

Contrary to the fact that the competitiveness of the firm is measured in terms of cost or performance measured directly on the current product in the long term. Thanks to this vision, this competitiveness is measured through the allocation of resources. These are unique, durable and difficult to imitate or substitute, hence the infrastructure implemented by the supplier. These infrastructures are linked to manufacturing resources (organization of manufacturing, human resources management, production planning, information system, etc.) or to technical resources (organization in research and development, staff qualification, investment Research laboratories and control equipment, relations with external experts, etc.). Thus, for each type of relation, the proposed model defines the criteria corresponding to the selection of the suppliers.

In conclusion, the selection of suppliers is a very complex process, which depends on several factors such as the company's sector of activity, the type of relationship to be undertaken between the company and

the suppliers, and so on. The various works in this field show, however, that the triptych QCD (Quality, Cost, Delay) remains the most used in this process.

These and other criteria are sometimes conflicting, making the process of supplier selection complicated. Several methods have been established for solving this problem. The following paragraph presents a state of the art of the main methods.

3 Presentation of the local integration project and diagnosis of the perimeter body & Electrical

3.1 Presentation of the local integration project and its role in improving purchasing performance

General presentation: Local integration of automotive parts:

This trend is motivated by the gains to be made in terms of geographic proximity, lower logistics costs and exchange rate risks. To benefit from the cost advantage, Renault Maroc is looking for local partners for business opportunities in order to establish a Win-Win relationship and also minimize logistical costs.

The supplier's competitiveness remains the main concern of Renault Maroc. To this end, it accompanies these suppliers with the ANPQP (Alliance New Product Quality, Procedure) approach, which is founded by Renault-Nissan to make annual agreements and provide specialized teams for suppliers to transfer know-how, especially in terms of quality management and competitive production.

The ANQP is the common procedure to cover all the quality assurance activities expected for the POE (Outsourced Product of the Renault plants). It was developed to define the common requirements of Renault and Nissan towards their suppliers from the initialization of the project, through the manufacturing agreement, to the end of production series. The challenge is therefore to reach the objectives qualities, cost, delivery time and overall expectations of the customer.

The problem that generally arises for the design of automobile parts in Morocco is the lack of technology and experience. Indeed, the Moroccan industrial fabric does not have the technologies necessary for the manufacture of high quality parts or the human skills necessary for the realization of a project of high posture. However, there are multinationals, established in Morocco to fill the gap, and already part of the Renault-Nissan supplier panel.

The criteria requested by Renault-Nissan Purchasing Organization (RNPO) is a joint purchasing organization between Renault and Nissan, which was set up in 2001 and has gradually evolved to cover 100% of purchases since 2009. Common purchases Mean that Renault and Nissan have mandated RNPO to build a business panel and monitor the performance of suppliers), for the choice of suppliers is grouped in 4 primary parts:

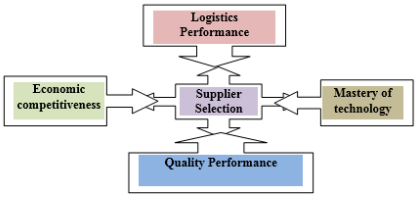


Figure 3: Supplier Choice Criteria at RNPO

Local integration is an effective tool for cost reduction, but the process is slow and requires heavy investment of tools and studies. A profitability study is carried out at the end of the consultations to decide on the profitability of the project for Renault.

Today, the rate of local integration is 33%, which remains very low compared to the targets set at the start of the Renault Tangier plant.

Indeed, the last supplier agreement signed in April 2016, and the establishment of ecosystems, will allow Renault to boost this rate of local integration thanks to permanent support from the Moroccan State, specifically the Ministry of Industry for The development of the industrial fabric in Morocco. The objective is to achieve a local integration rate of 65% by 2023, which remains a challenge for both stakeholders namely Renault and the Moroccan State. So the local integration project takes into account not only the optimization of costs but also the commitment vis-à-vis the government.

SWOT analysis of local integration:

The SWOT analysis of local integration will enable us to identify, on the one hand, the components on which we can rely and count, and on the other hand, the weaknesses that need to be reinforced. Here is the SWOT analysis:

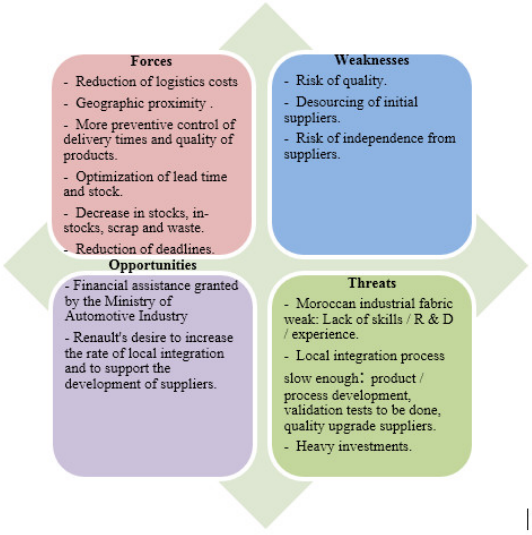


Figure 4: SWOT Analysis of Local Integration

Local integration steps:

The process of local integration takes place in several stages, from the determination of the perimeter impacted up to the appointment of the supplier, which makes sense, since sourcing is a flexible process that changes according to sector of activity and the nature of the business. Below the different actions carried out for each stage of local sourcing:

Step 1: Determine the scope of action:

In this phase, the following elements are specified:

- Identification of the parts concerned.
- Identification of projects.
- Identification of customer sites.
- Determination of current suppliers.
- Identification of diversity (All existing references), and calculate the mix of references by several criteria either by volume, power train, or product range, the choice of mix depends on the nature of the part.
- Location.
- Prospecting and determination of suppliers to consult (visiting local suppliers, consulting current foreign suppliers for a possible establishment in Morocco).

Step 2: Selection Criteria:

In this phase, we determine our supplier selection criteria, which are: Geo-location (Integrated multinational suppliers)

- Economic competitiveness.
- Technological mastery.
- Logistics performance.
- Quality performance.

Step 3: Define local integration scenarios:

After defining the selection criteria, it is recommended to draw a road map for possible scenarios of the local integration.

Step 4: Recovery of engineering definition:

Engineering prepares the technical definition in parallel with the purchase for 3D modeling and 2D of the part.

Step 5: The sending of consultation files to the supplier (Request for information):

The purchase sends the RFI file. This is the equivalent of the purchase specification.

Step 6: Kick-off meeting with suppliers:

The purpose of this meeting is to discuss the investments necessary for the start-up of the manufacture of the part

Step 7: Launching of logistics costing :

To calculate the necessary transport costs, and compare the new and the old logistic path.

Step 8: Comparative Analysis, Supplier Returns:

A comparison of the offer of new supplier and the offer of the current supplier, it is done on the following points:

- Price comparison.
- Comparison of logistics costs.
- Comparison TEI (Investment Ticket).
- Comparison of delivery times.
- Comparison of quality.

Step 9: Purchase Negotiation:

This stage of negotiation is essential for building a lasting relationship with the supplier.

Step 10: Appointment of Supplier:

The supplier is appointed for the delivery of the part.

Step 10: Appointment of Supplier:

The supplier starts the manufacturing process of the part.

Finally, this step is the standard procedure for a local sourcing project, the order of these steps differs from one project to another.

Before starting the project of our integration we will first analyze the inventory of the perimeter purchase Body & Electrical.

4 Analysis of the state of the premises and diagnosis of the perimeter body & Electrical:

Inventory of fixtures:

The starting point of the project is to do an analysis of the Body & Electircal perimeter of Renault Maroc. The perimeter that will be analyzed corresponds to everything related to the body of the car and the electronic elements and systems that constitute it.

The Body & Electircal perimeter is made up of 5 buyers, each one dealing with a number of parts and their suppliers. Some of these parts are locally integrated and another part is not. To do that, we tried to develop a supplier panel so that we could have more visibility.

In this perimeter there are 4 families of purchases, below a table that summarizes all the types of purchases in this perimeter:

Table 5: on the Body & Electrical perimeter parts panel

Body Equipment Purchasing	- Glazing. - Watertight seal. - Wing mirror. - Lighting. - Wiping.
Purchases climate and cooling system	- Air conditioner. - Radiators. - Radiator hoses. - Compressors. - GMV: Ventilation motor unit.
Purchasing Multi-Media and Mechatronics	- Clusters. - Door mechanism. - Hinge. - Switch. - Combi switch. - Door switches. - Electronic keypad. - Probes. - Speakers. - Antenna. - Microphone and displays. - Radio / Navigation system.
Buying Harness, Energy Systems	- Wiring. - Drums. - Starter.

Indeed, five buyers share among them these parts, their management, their supplier panel, and supply monitoring.

Indeed, this perimeter is very strategic, since it gathers a very interesting percentage of all parts worked on vehicles. That said, buyers must have background and training in the technical field and specifically, in mechanical and electronic engineering to handle negotiations very well with its supplier.

We have chosen the group of merchandise climate and cooling system, for a possible local integration, however before beginning the consultations, we have made a small diagnosis of this grouping of merchandise to study the parts not yet integrated locally and those potentially to be there in 2016.2016.

Diagnosis and analysis of the parts group Climate and Cooling System:

We have created a table that includes all the families of purchases and the parts allocated to them. Subsequently the document was sent to the buyer who is responsible for the perimeter climate and cooling, so that he can tell us which purchasing families he is responsible for, their suppliers solicited and the parts that are integrated locally and have not yet been.

This support will help us to have a visibility on the parts that can be integrated locally and to identify it in order to launch the supplier consultations on these parts.

Table 6: analysis of the climate and cooling system

The parts of the climate perimeter and cooling system		buyers	Local integration 2016	Suppliers	City	Integrated Potential 2017
Climate and cooling system	Compressors	X	No	Hutchinson	Spain	Yes
	HVAC manual		Yes	Denso	Tanger	
	HVAC regulated	Y	Yes	Denso	Tanger	
	GMV	Z	Yes	Suit/Gate	Tanger	
	Air guide		Yes	Sews	Kenitra	
	Radiator		Yes	GFD/Calsonic	Berrechid	
	Radiator hoses		No	Tristone/Superindus India e	Inde	Yes
	Hoses for water evacuation		Yes	MGI	El-Jadida	

This table is a summary on the integrated parts and the others that are potentially integrable for the year 2017 and for which we already have an idea on the feasibility of the project. This table contains the following:

- Suppliers solicited for each commodity group.
- The parts that will be integrated for the year 2016.
- Potentially integrable parts for the year 2017.

After drawing up this table, the engineering department and the project purchases made a meeting to choose between the compressors and the radiator pipes for a possible local integration. We found that there is no international supplier installed in Morocco that would have the necessary technology to produce compressors of good quality. In the end Renault chose to start on the local sourcing of radiator pipes.

5 The project framework

We concluded that MGI will be interested in the supply of radiator pipes as shown in the table above shows that Renault has already integrated locally at MGI El-Jadida. The water drain pipes that are more or less similar in their production than the radiator pipes, in addition MGI already has an experience effect in the production of radiator pipes since they produce this type of parts in its plants everywhere In the world for Renault and for other car manufacturers.

The choice of radiator pipes was validated by the engineering department, as well as the purchase projects including with the approval of Chamberlain Nicolas project manager local integration Renault Morocco. The supplier MGI with its new site in El-Jadida represents an opportunity to be seized by Renault. The table below summarizes the different levers of this project:

Table 7: Project Framework Summary

Target Supplier	MGI El-Jadida
Actual Supplier	Superindus India
Year of projection	2017/2018/2019
Turnover	300 000Euro
Projects	Lodgy, Sandero, Dacia Logan, Docker
Impacted perimeter	Radiator Inlet Hose Radiator Outlet Hose Heat Inlet Hose Hose outlet radiator heating Hose intake tank degassing Hose outlet tank degassing Hose purge

A precision, to clarify the confusion between the nature of the part and the references. Each part has a number of references, in the case of this project there are 800 references for the family of radiator pipe parts. The reference is a code which designates the technical characteristics of the part.

6 Consultation phase

Generally this phase is subdivided into two main components:

The development of the purchase specifications:

In the Renault standard the purchase specification is constituted in the form of an RFI, Request for information. Indeed for the realization of the mission it is necessary to collect different information from different departments so that the file is in good and due form.

The volume of vehicle projects:

At Renault each vehicle is associated with a specific code and a project manager responsible for the monitoring of the sourcing carried out. To do this consult the project managers purchase to have the volumes impacted by our sourcing on each type of vehicle namely: Docker, Dacia Logan, Sandero, lodgy.

Indeed, the designation of the volumes for the consultation is done according to a well-defined purchase strategy, to simplify, the project managers buy have two different volumes for each project, the first is

the average volume and the second c 'Is the maximum volume, this difference is due to the unpredictability of the scenarios throughout the duration of the project.

Generally, consultations are conducted on average volumes. This is what we did, for the simple reason that if the supplier is competitive on the average volume, it will automatically be competitive on the max volume.

Mix volumes and references:

The second information to be included in the RFI is the Mix volumes and references. The Mix volume is the criterion for which the supplier must return the volume of the previous table. In addition, since we do sourcing to radiator pipes, we must consider the types of motorization and the different mechanical aspects of the parts.

This information is retrieved from the engineering department. They provide us with the references of each type of radiator hoses and their mix with respect to the existing engines.

The references are like an engineering code that defines the degree of injection per second and the different mechanical dimensions. This information is strictly confidential. We were able to retrieve a table on the analysis of different types of motorization with their mixes in an anonymous way.

Quality Renault:

The purchasing function is a transversal function. This is reflected in the RFI's drafting, which must specify the quality standards that Renault requires from its supplier.

To do this, the quality department recovers the quality objective of Renault for the years (2016, 2017, 2018, 2019) of the radiator pipes, as well as the degree of incident tolerated by quarter. This last criterion is measured by what is called the PPM indicator. This is means incident by thousands.

- For example: a PPM = 5 goal, means that in a number of 1000 parts delivered to the Renault site, the degree of scrap and non-conforming parts must not exceed 5.

Indeed, Renault is developing another quality indicator for its suppliers. This is the ranking. It classifies its suppliers and awards a quality rating based on the experiences with the supplier, as well as its reputation and ISO quality certifications

After the recovery of the quality objectives, it is necessary to do the same for the logistics side. Except that the objectives of the logistics are the same as the previous year with the precision of the incoterm FCA. This precision gives transport responsibility to Renault.

The technical file:

The purchasing file consists of two main elements: the RFI and the technical file. The responsibility for the elaboration of this last one weighs on the shoulders of the engineers.

In fact, this file contains the technical definition of the parts, specifying the 2D and 3D models, and the technology used for quality production and consistent with Renault standards.

7 Analysis Phase

The following two indicators need to be analyzed:

- Transport encryptions.
- Current CA analysis.

Transport encryption:

In order to calculate the transport gains that Renault can generate from this project, it is necessary to carry out a transport calculation, comparing the current logistic flow and the future flow of the radiator pipes. Below is a diagram of the two flows:

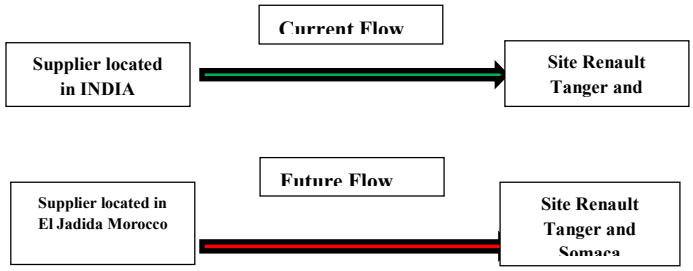


Figure 8: Logistic flows of radiator pipes

To compare the logistic costs of each stream there is a unit at Renault specializing in this transport encryption. It calculates the transport costs of the two flows.

The nature of the information collected for each reference is:

- The types of projects, in other words the different vehicles impacted.
- The weight of each reference: each reference has a specific weight. This information is provided by the engineering department of the company.

Daily rates: the volume that Renault wants to receive every day to avoid breaking the production chain.

- The code of the packaging unit (CPU): the type of CPU used in order to know the number of parts that this packaging unit can contain.
- The code of the handling unit (UM): the type of UM used in order to know the number of the CUs that the UM can contain.
- The supplier site account and the customer account.

After collecting this information from different departments, the launch of transport encryption for current flows will be possible.

After a kick off, the engineering department deducts the nature of the CUs and the UMs that the supplier must use since the transport is provided by Renault. Consequently, Renault launched the transport costing for this future flow.

The return of the two transport figures, made it possible to identify the transport gain that Renault can earn

The table below summarizes the results of the two streams. We cannot give all the details of transport gain for each reference because of confidentiality:

What we can deduce is that the transport gain will be 5.51 euro for each family of parts delivered to the Renault plant.

For this purpose the Purchasing Department has established a current CA analysis and the prices that Renault pays for the radiator hoses in order to compare with the supplier's offer.

Current turnover analysis:

The reference prices appear on a platform called SCOPP, which manages supplier contracts, their appointment, and the prices of all sourced references as well as supplier flows.

The list of references will be sent to a cell specializing in the administration of this platform, in order to retrieve the prices of each reference.

The objective was to realize the impact of total turnover of references consulted and to know the amount that Renault currently pays. Here is the result of the analysis in the form of a table:

Table 9: Turnover analysis

The parts/year	price
Radiator Inlet Hose	20 000,00 €
Radiator Outlet Hose	10 000,00 €
Heat Inlet Hose	30 000,00 €
Hose outlet radiator heating	40 000,00 €
Hose intake tank degassing	100 000,00 €
Hose outlet tank degassing	60 000,00 €
Hose purge	20 000,00 €
TOTAL	280 000,00 €

The turnover shown in this table is the amount that Renault is paying today. Vis-à-vis its supplier, it is necessary to specify that this table summarizes the turnover for each type of radiator pipes.

Indeed, each type of radiator pipes corresponds to a sum of references. At the end we have specified in the project framework that there are 800 references for all radiator pipe parts.

Supplier return

The supplier return was set at the second week of April, unfortunately, Renault did not receive the supplier offer on time.

As a result, Renault relaunched its demand, and this time Renault received an unfavorable return which expresses that the quantity of references sent is very large and it will take a long time to encrypt them on the form of an industrial process. In addition it has even specified that to produce the quantity requested it will need very large investments which are very expensive and that the requested technology does not exist in Morocco. Therefore Renault must expatriate it from their site located in Ukraine.

In short, Renault has deduced that the supplier is not competitive on all references.

8 Conclusion

Local sourcing is a very complicated process in non-industrialized countries. Thanks to a solid industrial fabric, an experience effect, and advanced technologies available within the country Renault was able to achieve a local integration rate of 90% in India.

The lack of technology and the effect of experience make it very difficult to establish local integration because investment costs are enormous. By default, the competitiveness of local suppliers in relation to international suppliers' prices will be impacted. Of course, ecosystem conventions have been established to solve this problem of non-competitiveness of local suppliers. This was possible through the subsidies that the Moroccan State undertakes to grant suppliers to encourage them to make the necessary investments.

Indeed, the problem of the departure was to know if the local sourcing always allows optimizing the purchase costs of the multinationals. To verify this hypothesis we have piloted a project of local integration for the multinational Renault.

We have found that local integration does not always optimize the purchasing costs of multinationals because of the lack of competitiveness of local suppliers, especially in developing countries.

REFERENCES

[1]. Op.cit., Bozarth, Handfield, Introduction to Operations and Supply Chain management, Pearson, 2007

[2]. Supply Chain Brain, U.S. and China Are the Riskiest Geographies for Sourcing and Manufacturing.

[3]. AberdeenGroup, Low-Cost Country Sourcing : Success Strategies Maximizing and Sustaining the Next Big Supply Savings Opportunity, Juin 2005

[4]. Sara Ireton, Global Sourcing Checklist. Industry Week, janvier 2007, p. 40

[5]. Mike Jones, Domestic costs of foreign sourcing. Material Handling Management, février 2008, p. 40-41

[6]. Sean Delaney, Delivering Value in both good and bad times, E-sourcing Forum, 14 juillet 2008

[7]. N. Sheth, « présentation d'un modèle de comportement des acheteurs industriels, Le Supplier Choice Model » WOODSID & VYAS (1984;1986; 1987)

[8]. Roger Pérotin marketing achats, stratégies et tactiques, Edition 2010

[9]. SidhuSephardi, Sourcing strategy, principles, policy and designs, 2005, Springer

[10]. Salons Solutions, discussion informelle avec des décideurs, CNIT Paris La Défense, 30 septembre 2008.

[11]. Collectif, Encyclopédie de gestion, article « Stratégie d'entreprise et économie industrielle », écrit par A. Desreumaux, Economica, 1997.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Contextual Arabic Handwriting Recognition System Using Embedded Training Based Hybrid HMM/MLP Models

[1]Mouhcine Rabi, [2]Mustapha Amrouch, [3]Zouhir Mahani
[1,2]Laboratory IRF-SIC, faculty of sciences IbnZohr University, Agadir Morocco
[3]High school of technology, IbnZohr University, Agadir Morocco
mouhcineh@gmail.com, m.amrouch@uiz.ac.ma, zouhir.mahani@uiz.ac.ma

ABSTRACT

Recognizing unconstrained cursive Arabic handwritten text is a very challenging task the use of hybrid classification to take advantage of the strong modeling of Hidden Markov Models (HMM) and the large capacity of discrimination related to Multilayer Perceptron (MLP) is a very important component in recognition systems.The proposed work reports an effective method on improvement our previous work that takes into consideration the context of character by applying an embedded training based HMMs this HMM is enhanced by anArtificial neural network that are incorporated into the process of classification to estimate the emission probabilities. The experiments are done on the same benchmark IFN/ENIT database of our previous work to compare the results and show the effectiveness of hybrid classifier for enhancing the recognition rate the results are promising and encouraging.

Keywords : Arabic Handwriting Recognition; Context; Embedded training; HMMs; Multilayer Perceptron (MLP).

1 Introduction

Systems for handwriting recognition are referred to as off-line or on-line systems depending on whether ordinary handwriting on paper is scanned and digitized or a special stylus and a pressure-sensitive tablet are used. In both the ultimate objective is to convert handwritten sentences or phrases words or characters in analogue form (off-line or on-line sources) into digital form (ASCII).

More than 300 million people around the world speak Arabic and their derivative. Arabic is naturally written cursively in both handwritten and typewritten modes. In comparison to Latin Arabic seem to be more complex. For example many letters in this language have complementary diacritics such as dots madda and zigzag bars. In addition the letters have different shapes at different locations of the word.

Due to variability in handwriting styles and distortions caused by the digitizing process even the best handwritten word recognizer is unreliable when the number of word choices is large. This necessitates the use of advanced concepts to achieve a performance level comparable to that of humans. The researches focus on the use of new methods and approaches to perform handwriting recognition. In this area the concept of combining multiple classifiers is proposed as a new direction for the development of highly reliable handwriting recognition systems and some preliminary results have indicated that the

DOI: 10.14738/tmlai.54.2983
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2983

combination of several complementary classifiers will improve the performance of individual classifiers[01] [02].

In current work our system performs training and recognition of words and characters. In order to model the variations related to the character context in the corpus we have opted for a specific type of learning. Therefore character models are trained and obtained through embedded training; thereafter the decision has been done by the proposed hybrid classifier which mainly based on HMMs and neural network.

The remainder of this paper is organized as follow. Section 2 presents a literature review and related works of handwriting recognition system. Section 3 is focused on our contribution starting with our developed reference system then the incorporation of hybrid classifier. The performance of the recognition system has been experimented on the benchmark database IFN/ENIT and the obtained experimental results are shown and analysed in section 4. The paper finally closed with a conclusion and perspectives.

2 Literature review

Hidden Markov Models (HMMs) have proven to be one of the most successful and widely used classifiers in the area of text recognition. There are many reasons for success of HMMs in text recognition including avoidance of the need to explicitly segment the text into recognition units; characters or graphemes In addition HMMs have sound mathematical and theoretical foundation [03].

[04]The authors have investigatedon contextual sub-characters HMMs for text recognition by using multi-stream HMMs where the features calculated from a sliding window frame form one stream and its derivative features are part of the second stream. The experiments were conducted with different train-test configurations on the IFN/ENIT database and the best recognition rate achieved was 85.12%. In [05]Azeemused an effective technique for the recognition of offline Arabic handwritten words using Hidden Markov Models. Besides the vertical sliding window two slanted sliding windows are used to extract the features. Three different HMMs are used: one for the vertical sliding window and two for slanted windows a fusion scheme is used to combine the three HMMs. [06] proposed a combined scheme for Arabic handwritten word recognition using a HMM classifier followed by re-ranking. Basically intensity features are used to train the HMM and topological features are used for re-ranking to improve accuracy. Experiments were carried out using IFN/ENIT database and the achieved recognition rate was 83.55%.An alternative approach is proposed in [07] and [08] where multi-stream HMM models is used this paradigm provides an interesting framework for the integration of multiple source of information. Significant experiments have been carried out on two public available database the recognition rate was 79.8% in IFN/ENIT for Arabic and 89.8% in IRONFF for Latin script. In [08] Maqqorproposed a system of Arabic handwriting recognition based on combining methods of decision fusion approach. The combination of the multiple HMMs classifiers was applied by using the different methods of decision fusion approach. The system is evaluated using the IFN/ENIT database. Experimental results demonstrate that the Weighted Majority Voting (WMV) combination method have given better recognition rate 76.54% with Gaussian distribution. In [09] the authors present an analytical approach of an offline handwritten Arabic text recognition system. It is based on the Hidden Markov Models (HMM) Toolkit (HTK) without explicit segmentation.The feature extraction uses a sliding window on the line text image and processed by two groups of these features (the features of local densities and the statistical characteristics).The proposed

system has been experimented in two different databases: “Arabic-Numbers” database where we achieved a rate of 80.26% for words and 37.93% for sentences and IFN/ENIT database where we achieved a rate of 78.95% for words.

Otherwise [10] proposed a new approach for the offline Arabic handwritten word recognition based on the Dynamic Hierarchical Bayesian Network(DHBN) using a free segmentation released by a smoothed vertical projection histogram with different width values. The model is consisting of three levels. The first level represents the layer of the hidden node which models the character class. The second layer models a frame set representing the sub-characters and the third layer models the observation nodes. The developed system has been experimented and the results are provided on a subset of the IFN/ENIT benchmark data base. These results show a significant improvement in the recognition rate because of the use of the DHBN. Most of the recognition errors of the proposed system can be attributed to the segmentation process error and to the poor quality of some data samples.

In many other works neural networks in its different applications have been extensively applied to classify characters as part of isolated or continuous handwritten word recognizers [11] [12] [13] [14] [15].

This paper focus on the impact of using aembedded training based on a hybrid classifier the motivation for the work on the hybrid HMMs and Artificial neural network models presented here originates from a critical analysis of the state of the art in offline handwritten text recognition[16] [17] [18] our previous work on offline handwriting recognition using HMMs [19] researches and experiences in using hybrid HMM/ANN models for automatic speech recognition [20][21][22][23] [24]and for online handwriting recognition [25]. All these criteria making hybrid modeling an important factor in order to achieve aneffective and efficient system.

3 Contribution

3.1 Reference System [19]

Our reference system (figure1)was analytical without explicit segmentation based HMMs using embedded training to perform and enhance the character models.

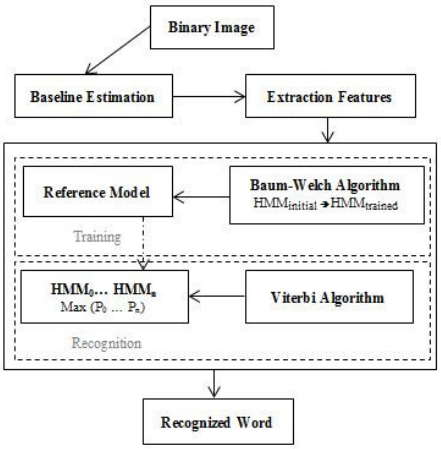


Figure.1. Synopsis of reference system

Extraction features was preceded by baseline estimation; the approach used to estimate these baseline based on the horizontal projection curve that is computed with respect to the horizontal pixel density

knowing that the skew and slant correction of words are made in pre-processing step to harmonize the direction of the sliding windows in the extraction features. These latter are statistical and geometric to integrate both the peculiarities of the text and the pixel distribution characteristics in the word image.

The sliding windows are shifted in the direction of writing (right to left). In each window we extract a set of 28 features represent the distribution features based on foreground pixels densities and concavity features. Each window is divided into a fixed number n of cells. Some of these features are extracted from specific areas of the image delimited by the word baselines.

These features are modelled using hidden Markov models and trained by theembedded training method (figure 2).

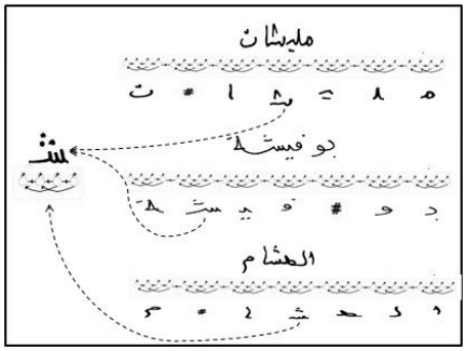


Figure.2. Embedded training of character "chin" [19]

We used a model for each character right-left topology with four states and three transitions for each state. Word model is built by concatenating the appropriate character models.

The embedded training is to automatically identify relevant information letters without specifying them explicitly by exploiting the redundancy of information between words matched to changes in context and letters position.

The major problem of HMMs is the estimation of emission probabilities; this confirms that HMMs are powerful to model sequences but still limited compared to NN and SVM in classification [26] for this reason and the motivations cited above in section II it is prominent and promising to use a hybrid classifier.

For more explanations about the baseline system refer to [19].

3.2 Hybrid Classifier

To improve the performance of the off-line handwritten recognition system either the accuracy of the classifier has to be increased. In this section we introduce the hybrid approach then we clarify the principle for our offline Arabic handwriting recognition system.

While HMMs are effective in modeling variation in handwriting they lack discrimination ability because of maximum-likelihood parameter estimation criteria. The strength of MLP is in the fact that they don't need to assume about statistical distribution of input as wellas they can be trained to exhibit discriminant properties. As already mentioned recent works in various area of research tried to develop hybrid HMM/MLP systems in which MLPs areused to compute the emission probability associated with each state of HMM.

By HMM the goal of handwriting recognition is to retrieve the most likely grapheme character or word sequenceW given a sequence of observation vectors x which is achieved by maximizing thea posteriori probability:

$$\hat{w} = \arg\max P(w \mid X) \quad (1)$$

Typically each grapheme is modelled by a right to left HMM and the number of states is chosenglobally or individually for each character. Gaussianmixtures are used to model the output distributions ineach state q given the feature vector xP(x|q). The Baum-Welch algorithm is used for training the HMMs whereasthe Viterbi algorithm is used for recognition.

Hybrid modelsfor handwriting recognition based HMMs were built with different neural networks.[27]; [28] use an MLP. [29]built an hybrid CNN/HMM. In [30]CNNs is applied inthe hybrid framework for handwritten word recognition using different segmentationmethods.

In Hybrid HMM/MLP classifier neural networks can be considered statistical classifiers under certain conditions by supplying output of a posteriori probabilities. Thus it is interesting to combine the respective capacities of the HMM and the MLP for a new efficient recognition system inspired by the two formalisms.

The principal idea behind the MLPNN/HMM hybrid approach as illustrated in Figure 3 is to estimate the output probabilitydensity function of each state of the used HMM by the output nodes of the MLP classifier which received features as input. These input vectors are pre-processed to finally estimate the posteriori probability deciding whether the input vector belongs to the desired character class. The MLPNN's output weighted by the priori probability of each class forms the probability density function used for every state of the HMM.

In the hidden Markov modeling approach the emission probability density P(x|q) must be estimated for each state q of the Markov chains that is the probability of the observed feature vector x given the hypothesized state q of the model.

In the proposed hybrid HMM/ANN approach the emission probabilities are provided with a neural network since ANNs can be trained to estimate probabilities that are related to these emission probabilities. In particular an MLP can be trained to approximate the a posteriori probabilities of states P(q|x) if each MLP output unit is associated with a specific state of the model and if it is trained as a classifier.

The a posteriori probability estimates from the MLPoutputs P(q|x) can be converted to emission probabilitiesP(x|q) by applying Bayes rule:

$$\mathrm{P}(x \mid \mathrm{q}) = \frac{P(q \mid x)P(x)}{P(q)} \quad (2)$$

The class priors P(q) can be estimated from the relativefrequencies of each state from the information produced bya forced Viterbi alignment of the training data. Thus thescaled likelihoods P(x|q)/P(q) can be used as emissionprobabilities in the proposed system since during recognitionthe scaling factor P(x) is a constant for all classes.This allows MLPs to be integrated into hybrid structuralconnectionistmodels via a statistical framework.

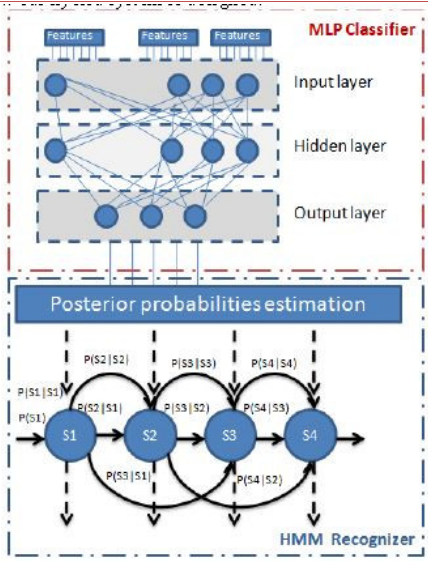


Figure 3: Global scheme design of the hybrid model HMM/MLP

The advantages of this approach are the discriminate training criterion (all MLP parameters are updated in response to every input feature vector) and the fact that it is no longer necessary to assume an a priori distribution of the data. Furthermore if left and right contexts are used at the input of the MLP important contextual information can be incorporated into the probability estimation process.Another strength of this approach is that computing emission probabilities with hybrid HMM/ANN models is usually faster than conventional HMMs with Gaussian emissions since it only requires a forward pass of the MLP for all states of the Markov chains.

4 Experimental results

To evaluate the performance of our recognition system experiments are conducted using IFN/ENIT [31] database of handwritten Arabic words. It was produced by the Institute for Communications Technology at the Technical University of Braunschweig (IFN) and the “Ecole Nationale d’ingénieur de Tunis (ENIT) “National school of engineering Tunis.”This database is used bymore than 110 research groups in about 35 countries [32].

4.1 HMM vs HMM/MLPClassification

Reference recognition HMM experiments were conducted using continuous density HMMs with diagonal covariance matrices of Gaussians in each state, a right-to-left topology was applied with four states for each character and three transitions for each state. The optical models were trained and tested using the HTK toolkit [33].

The developed system used hybrid classifier based HMMs enhanced by anartificial neural network that are incorporated into the process of classification to estimate the emission probabilities.

Table1 shows the experimental results of our developed system compared to reference system using the same benchmarking database IFN/ENIT to illustrate the reliability of ourimprovement models.

Table1: Recognition results of improvement system compared to reference system

System	Models	RR* %
Reference	HMM	87.93
Hybrid	HMM/MLP	**89.03**

RR: Recognition Rate

We compare the reference system [19] with the developed system using hybrid HMM/MLP classifier. Results in Table1 show an improvement due to embedded training using thehydride classification: accuracy is increased by 1.1%.

4.2 Comparison with other systems

A comparison of the recognition rates of our system with other state-of-the-art systems evaluated on the IFN/ENIT database is presented.

Table2 shows the results of recognition rates for various offline systems recognition of cursive Arabic handwritten text using diverstype of models and the same database with the same configuration (training-testing); sets a, b and c for training and d for test, to compare rates and infer the effectiveness of the proposed method.

Table 2. Recognition results of various systems

System	Models	RR* %
Irfan [04]	Contextual sub character	85.12
Alkhateeb[06]	HMM+ re-ranking	83.55
Kessentini[07]	Multi-stream HMM	79.80
Maqqor[08]	Multiple classifier	76.54
ElMoubtahij[09]	HMM	78.95
Khaoula[10]	DBN	82.00
Our system	**HMM/MLP**	**89.03**

As it can be noted from Table 2 most of theprevious systems are basedon HMM using various techniques exploiting the contextual approach, multiple HMM classifier, re-ranking or other techniques to improve HMM models and the recognition rate for the results of the systems mentioned does not exceed 86%. Others used hybrid classifier such as HMM and Dynamic Bayesian Network and the recognition rate achieved was 82.00%. Whereas the proposed system using HMM/MLP outperforms the results and achieve 89.03% due to enhancement of HMMs by incorporating anartificial neural network into the process of classification to estimate the emission probabilities..This illustrates the effectiveness of embedded trainingto take account the context of characters to perform the models and using hybrid HMM/MLP classifier to improve the performance of recognition system.

5 Conclusion and perspectives

In this paper we have enhanced the HMM based reference system by using a hybrid HMM/MLP classifier.Extracted features are statistical and geometric to integrate both the peculiarities of the text and

the pixel distribution characteristics in the word image. These features are modelled using hidden Markov models. These models as already mentioned in [19] take into account the context of character by applying aembedded training to perform the models. In addition, the contribution in this paper is the improving of HMM modeling by incorporating MLPs to estimate emission probabilities that present the major HMM problem in order to take advantage of the strength of HMM modeling and neural networks classification.The modelling proposed has improved recognition and shown encouraging results to be perfect using annexes improvements.

Due to variability in handwriting styles even the best handwritten word recognizer is unreliable when the number of word choices is large. This forced the use of linguistic constraints to enhance HMMs modeling by a statistical language model that are incorporated as a post-processing into the process of recognition.

Statistical Language Modeling involves attempts to capture regularities of natural language in order to improve the performance of various natural language applications, e.g. , Speech recognition, Machine translation, Handwriting recognition, Information retrieval and other applications.

The goal of Statistical Language Modeling is to build a statistical language model that can estimate the distribution of natural language as accurate as possible, which could improve significantly the results especially when we extend our system for line and paragraph recognition.

REFERENCES

[1] J. H. Alkhateeb O. Pauplin J. Ren J. Jiang Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition In Knowledge-Based Systems 24 (2011) pp.680-688.

[2] Salvador Espan˜a-Boquera Maria Jose Castro-Bleda Jorge Gorbe-Moya and Francisco Zamora- Martinez “Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN models” 2011.

[3] T.Plots and G.A.Fink; “ Markov models for offline handwriting recognition: a survey” IntJ.Anal.Recongitvol 12 no 4 pp 269-298 2009.

[4] Irfan Ahmad GernotA.Fink and SabriA.Mahmoud “Improvement in Sub-character HMM Model Based Arabic Text Recognition” 2014 14th International Conference on frontiers in Handwrting Recognition.

[5] S. Azeem and H. Ahmed “Effective technique for the recognition of offline Arabic handwritten words using hidden Markov models” Int. J. Doc. Anal.Recognit.vol. 16 no. 4 pp.399–412 2013.

[6] J. AlKhateeb H. RenJinchang Jiang Jianmin Al-MuhtasebHusni Offline handwritten arabic cursive text recognition using hidden markov models and re-ranking Pattern Recognition Lett. 32 (8) 1081–1088 2011.

[7] Y. Kessentini T. Paquet A. Ben Hamadou Off-line handwritten word recognition using multistream hidden Markov models in: Pattern Recognition Letters 31 (2010) pp. 60-70

[8] A. Maqqor A. Halli K. Satori and H. Tairi Off-line recognition Handwriting combination of mutiple classifiers In 3rd International IEEE Colloquium on Information Science and Technology IEEE CIST'14 October 2014.

[9] El MoubtahijHichamHalliAkram Khalid Satori "Using features of local densities statistics and HMM toolkit (HTK) for offline Arabic handwriting text recognition" 2016

[10] KhaoulaJayech Mohamed Ali Mahjoub and Najoua Ben Amara "Arabic Handwritten Word Recognition based on Dynamic Bayesian Network" 2016.

[11] Berend-Jan van der Zwaag "Handwritten Digit Recognition: A Neural Network Demo" 2016

[12] Xu Chen Convolution Neural Networks for Chinese Handwriting Recognition 2016.

[13] Charlie Tsai "Recognizing Handwritten Japanese Characters Using Deep Convolutional Neural Networks" 2016

[14] ThéodoreBluche "Deep Neural Networks for Large Vocabulary Handwritten Text Recognition" 2015

[15] Ahmed Mahdi ObaidHazem M. El Bakry M.A. Eldosuky IVA.I.Shehab "Handwritten Text Recognition System Based on Neural Network" 2016

[16] AL-Shatnawi M. Atallah AL-SalaimehSafwan AL-Zawaideh Farah Hanna Omar Khairuddin"Offline arabic text recognition an overview" World Comput. Sci. Inform. Technol. J. 1 (5) 184–192 2011.

[17] M.T Parvez and S.A Mahmoud " Offline Arabic handwritten text recognition : A survey " ACM ComputSurvvol 45 no 2 pp 23-35 May 2013.

[18] A.Lawgali : "A Survey on Arabic Character Recognition" International Journal of Signal Processing Image Processing and Pattern Recognition. Vol. 8 No. 2 (2015) pp. 401-426.

[19] M.RabiM.AmrouchZ.MahaniD.Mammass "Recognition of cursive Arabic handwritten text using embedded training based on HMMs" Published in: Engineering & MIS (ICEMIS) International Conference on Sept. 2016 INSPEC Accession Number: 16467172 DOI: 10.1109/ICEMIS.2016.7745330 Publisher: IEEE

[20] Mohamed ETT AOUIL Mohamed LAZAAR Zakariae EN-NAIMANI "A hybrid ANN/HMM models for arabic speech recognition using optimal codebook"8th International Conference on Intelligent Systems: Theories and Applications (SITA), IEEE 2013.

[21] Supriya S. Surwade "Speech Recognition Using HMM/ANN Hybrid Model"International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169Volume: 3 Issue: 6 4154 – 4157.2015

[22] A. I. G-Moral, U. S-Urena, C. P-Moreno and F. D-Maria, 2011, "Data balancing for efficient training of hybrid ANN/HMM automatic speech recognition, " IEEE trans. on audio, speech and lang. proc., Vol 19, No. 3, 468-481.

[23] Anuj Mohamed K.N. Ramachandran Nair "HMM/ANN hybrid model for continuous Malayalam speech recognition"Published by Elsevier Ltd. Selection and/or peer-review under responsibility of ICCTSD 2012(International Conference on Communication Technology and System Design).

[24] E. Trentin, M. Gori, 2003, "Robust combination of neural networks and Hidden Markov Models for speech recognition," IEEE trans. on Neural net., Vol 14 , No. 6, 1519-1531.

[25] NajibaTagouguiHoucineBoubakerMonjiKherallah Adel M. ALIMI "A Hybrid NN/HMM Modeling Technique for Online Arabic Handwriting Recognition" International Journal of Computational Linguistics Research Volume 4 Nimber 3 Septembre 2013 pp 107-118.

[26] M.RabiM.AmrouchZ.MahaniD.Mammass" Evaluation of Features Extraction and Classification Techniques for Offline Handwritten Tifinagh Recognition " Global Journal of Computer Science and Technology (USA),: C Software & Data Engineering, Volume 16 Issue 5 Version 1.0 Year 2016.

[27] Dreuw P. Doetsch P. Plahl C. Ney H. (2011). Hierarchical hybrid MLP/HMM or rather MLP features for a discriminatively trained Gaussian HMM: A comparison for offline handwriting recognition Image Processing (ICIP) 18th IEEE International Conference on p. 3541-3544 Sept.

[28] S Espana-Boquera MJ Castro-Bleda J Gorbe-Moya F Zamora-Martinez "Improving offline handwritten text recognition with hybrid HMM/ANN models" IEEE transactions on pattern analysis and machine intelligence 33 (4) 767-779

[29] Q.GuoF.WangJ.LeiD.TuG.Li "Convolutional feature learning and Hybrid CNN-HMM for scene number recognition" Published in journal Neucomputing archive Volume 184 Issue C April 2016 Pages 78-90 Elsevier Science Publishers B.V Amsterdam The Netherlands.

[30] Bluche T. Ney H. &Kermorvant C. (2013b).Tandem HMM with convolutional neural network for handwritten word recognition.In 17th International Conference on Acoustics Speech and Signal Processing (ICASSP) (pp. 2390--2394).IEEE.

[31] M. Pechwitz, S.S. Maddouri, V. Maergner, N. Ellouze, H. Amiri IFN/ENIT – Database of Handwritten Arabic WordsCIFED 2002, Hammamet, Tunisia (2002), pp. 129–136

[32] Märgner V. El Abed H.: ICDAR 2011 - Arabic handwriting recognition competition. In: Int'l Conf. Document Analysis and Recognition pp. 1444–1448 (2011)

[33] S. Young al. the HTK Book V3.4 Cambridge University Press Cambridge UK 2006

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Mongo2SPARQL: Automatic and Semantic Query Conversion of MongoDB Query Language to SPARQL

Nassima Soussi, Abdeljalil Boumlik, Mohamed Bahaj
Department of Mathematics & Computer Science, Faculty of Science and Technologies, Hassan 1st University, Settat, Morocco
nassima.soussi@gmail.com, boumlik.abdeljalil@gmail.com, mohamedbahaj@gmail.com

ABSTRAT

In the last decades, the web has experienced a quantitative explosion of digital data handled by companies or organizations, prompting web users to switch to NoSQL system dedicated to Big Data in order to support large web sites destined for a very big audience due to the scalability and high availability of this system. In the other hand, the semantic web technologies has emerged with their considerable performance in data management by giving the web information a well-defined meaning, and allowing machines to intelligently access to different data sources. However, there is no bridge or an open extension toward these two systems in order to facilitate the interconnection between them; in addition, each NoSQL database has its own query language and does not support the standards of other systems (such as semantic web). All these reasons have motivated us to operate in this topic in order to unify the NoSQL query language and contribute in the interoperability of the both world with a specific focus on MongoDB as a NoSQL document oriented database by proposing the first provably semantic preserving algorithm named Mongo2SPARQL which transform MongoDB query language queries to its equivalent SPARQL ones taking into account key/value pairs having value's type array and embedded document, generating lists/bags and blank nodes after their transformation.

Keywords: Big Data; Mongo2SPARQL; MongoDB; NOSQL; Semantic Web; SPARQL.

1 Introduction

Over the past few years, the volume of data handled on the web is increasing day after day with a dizzying speed. In addition, the increasing computerization of this data processing implies an exponential multiplication of the data volume that can come from various sources such as: social networks, medical databases, telephone operators, economic data, scientific data, national agencies of territory defense and others; this big quantity of data makes its management and processing a real challenge that overtakes capacities of the traditional and classic databases engines. All these reasons led web users to switch to other technologies offering more robust database management systems dedicated specially for Big Data; we quote as example NoSQL and Semantic Web which represent the subject of this paper.

Regarding the oldest system that is semantic web [10], it represent an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation; it is based on the standards and protocols of the current web (http, URI and XML) and its

DOI: 10.14738/tmlai.54.2984
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2984

own standards: The Resource Description Framework RDF [11] dedicated to describe data, the Web Ontology Language OWL [12] for creating structured ontologies and the query language SPARQL [2] for querying data from RDF graphs. This system aims to exploit its semantic concepts in order to enhance big data technologies and help businesses to make the better decisions in real time. On the other hand, the second system named NoSQL (Not Only SQL) is a no relational database management system dedicated to manage heterogeneous and unstructured data. This system avoids join operations and supports dynamic schema design offering to web users a high flexibility and scalability. We distinguish four categories of NoSQL databases [1] with different architectural characteristics: key-value databases, graph databases, column databases and document databases with a specific focus on MongoDB. Since this two systems aim to make the big data processing smarter, establishing a connection between them is a very relevant need.

The remainder of this paper is structured as follow: Section II exposes some existing related works in this topic. Section III introduces the context of this work by presenting the architecture of our proposing method named Mongo2SPARQL and then it describes by detail each component of the previous architecture. Section IV exposes the application that we have developed so as to materialize our work. Finally, section V concludes this work and suggests some future extensions of our approach.

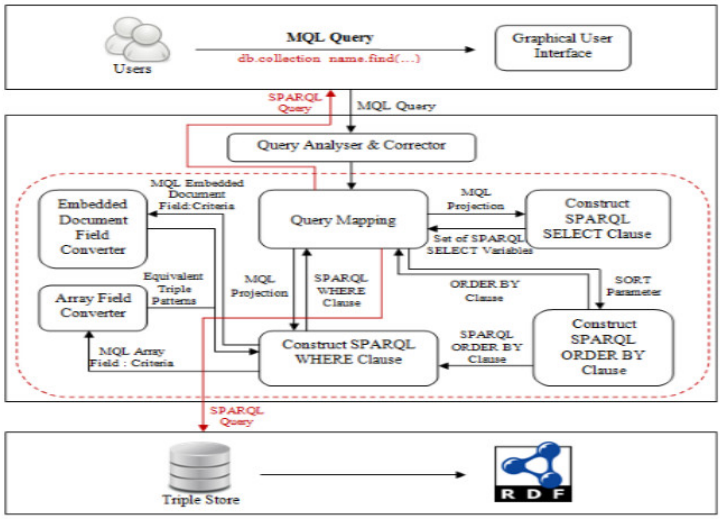


Figure 1.Mongo2SPARQL Architecture

2 Related Works

Recently, several researches focus on mapping data, models, concepts, and queries from the existing data source content to NoSQL world. The majority of these researches are interested much more to the relational systems than others; many approaches have been proposed about this mapping direction, such as: NoSQLayer [7] and others [8, 9].

Regarding the semantic web data source which represents the subject of this paper, all approaches proposed in order to bridge the gap between this latter and NoSQL database (MongoDB) are interested much more to the mapping direction of semantic web to MongoDB; we quote for instance: the paper [5] that proposes a query translation solution from SPARQL to query languages of the NoSQL sources

considered in this work (MongoDB as a document database and Cassandra as a column family store). This transformation is realized based on an intermediate query language named Bridge Query Language (BQL). The authors in [4] have proposed a basic approach that aims to handle document-oriented database as a triple store; they proposed an idea of mapping algorithm that translate SPARQL to Mongo Query Language taking into account single/multiple/optional pattern(s), literals, regex function of regular expressions, numeric values and distinct/alternative matching. The work described in [6] proposes a method to access arbitrary MongoDB JSON documents with SPARQL using custom mappings described in the xR2RML mapping language.

Therefore, to the best of our knowledge, this paper is the first work which investigates the reverse direction i.e. to look at RDF through NoSQL lenses by developing an efficient algorithm transforming MongoDB queries to SPARQL ones semantically equivalents in order to facilitate the triple stores data access to NoSQL users.

3 Research Architecture and Algorithm

3.1 Architecture Design

The proposed architecture as shown in Fig. 1 illustrates the different steps that guarantee the interoperability between MongoDB and semantic world. Our system is composed of seven components:

- Query Analyser & Corrector,
- Construct SPARQL SELECT clause,
- Construct SPARQL WHERE clause,
- Embedded Document Field Converter,
- Array field Converter,
- Construct SPARQL ORDER BY clause
- Query Mapping.

3.2 Architecture's Components Description

In this section, we will describe the different components composing our query mapping architecture designed in the paragraph bellow:

(1) Query Analyser and Corrector: this component aim to scan and analyze MQL query so as to check syntactic errors and correct them before starting the mapping process.

(2) Construct SPARQL SELECT clause: it takes as input a MongoDB query projection indicating the fields which will appear in the query result. Firstly, It glances through the projection list in order to extract from each element the field

label and field parameter, and then it verify the parameter's value that indicates the presence (if it is equal to 1) or the absence (if it is equal to 0) of this field in the SELECT clause of the equivalent SPARQL query. If the projection variable is NULL then the equivalent SELECT clause is (*).

```
Input: A MongoDB query projection, P
Output: A SPARQL SELECT Clause
Begin
select = " "
if ( P = NULL) then
    select = "*"
else
    for (i = 0; i < P.size(); i++) do
      field = P[i].getField();
      field_param = P[i].getFieldParameter();
      if (field_ param == 1) then
          select += "?" + field
      end if
      if ( i < P.size()-1 ) then
          select += " , "
      end if
    end for
enf if
return select
End Algorithm
```

3.3 Construct SPARQL WHERE clause:

This component takes as input a MongoDB query criteria so as to return at the end the SPARQL WHERE clause. Firstly, the algorithm glances through this list of criteria so as to extract from each element the field label and field criteria, and then it defines the subject of this iteration. Next it constructs the triple(s) pattern(s) and filter constraint by checking the type of the current field; if the type is ArrayField then our algorithm uses the component Array Field Converter so as to construct the equivalent WHERE Clause, else if the type is EmbeddedDocField then we use the component Embedded Document Field Converter, finally the last case is verified if the field type is SimpleField. Indeed, in this last case, we construct at first the triple pattern by concatenating the current subject, predicate and object of the current iteration, then we verify if the field criteria's type is a value so the filter constraint expression is written as Object = Field Criteria, else if the field criteria's type is a condition then the filter constraint expression is written in the form of Object Op Value with Op denotes Operator such as =, !=, >, >=, <, <=. Still remaining in the loop, this component verifies if the filterConst variable (denotes Filter Constraint) is not empty so as to treat the SPARQL FILTER clause. At the end, this previous results is concatenating in order to form a WHERE

clause constituted of a basic graph pattern only if the filterConst variable is empty or a basic graph pattern and filter clause in the opposite case.

```
Input: A MongoDB query criteria, C
Output: A SPARQL WHERE Clause, where
Begin
where, TP, subject, predicate, object, filterConst, field, field_criteria = "";
filter = "FILTER("; cpt = 0;
for (i = 0; i < C.size(); i++) do
   field = C[i].getField();
   field_criteria = C[i].getFieldCriteria();
   //define the subject
   if (field = "_id") then
      subject = field_criteria;
   else
     subject = "?s";
   end if
   // Construct SPARQLWhere Clause components
   if (field.type() = ArrayField) then
          Awhere = ArrayFieldConverter();
       TP += Awhere['TP'];
       filterConst += Awhere['filterConst'];
   else if (field.type() = EmbeddedDocField) then
          EDwhere = EmbeddedDocumentFieldConverter();
          TP += EDwhere['TP'];
          filterConst += EDwhere['filterConst'];
   else if (field.type() = simpleField) then
          predicate= field; object = "?"+field
          TP += subject + "" + predicate + "" + object
          if (field_criteria.isValue() = true) then
           filterConst = object + "=" + field_criteria;
       else if (field_criteria.isCondition() = true) then
```

```
            operator = field_criteria.operator();
            value = field_criteria.value();
        if (operator = "$gt") then
            filterConst = object + ">" + value;
        else if (operator = "$gte") then
            filterConst = object + ">=" + value;
        else if (operator = "$lt") then
            filterConst = object + "<" + value;
        else if (operator = "$lte") then
            filterConst = object + "<=" + value;
        else if (operator = "$eq") then
            filterConst = object + "=" + value;
        else if (operator = "$ne") then
            filterConst = object + "!=" + value;
        end if
   end if
end if
 if (filterConst.isEmpty() = false) then
     cpt++;
     if (cpt = 1) then
         filter += filterConst;
     else if (cpt>1) then
         if (C[i].type() = "field_criteria" OR C[i].type() = "and_criteria") then
                 filter += "&&" + filterConst;
         else if (C[i].type() = "or_criteria") then
                 filter += "||" + filterConst;
         end if
     end if
 filterConst = "";
 end if
```

```
end for
if (filter.isEmpty() = true) then
    where = TP ;
else
    where = TP + filter + ")" ;
end if
return where
End Algorithm
```

3.4 Array Field Converter:

This component takes as input an array type field, its criteria and its relative subject. This algorithm construct the triple(s) pattern(s) and filter constraint(s) equivalents to the input key/value (field/field_criteria) depending on the field criteria type if it is a simple value or an embedded document. It returns at the end an associative array of SPARQL Where Clause elements; the first element is a concatenation of triple(s) pattern(s) obtained and the second is a filter constraint(s).

```
Input: An array type field, field; Criteria of this field, field_criteria; subject
Output: An associative array of SPARQL Where Clause elements of the input field, Awhere[],
Begin
TP, S, P, O, fieldConst = ""; Awhere[]; L=0;
if (field_criteria.type() = Array) then
    S=subject; P= field; O= "?x"
    TP += S + "" + P + "" + O
    L = field_criteria.size();
    for (i=1; i<=L; i++)
 //Construct an equivalent Triple Pattern of the current criteria
          S ="?x"; P= "rdf:_"+ i ; O= "?o"+ i
          TP += S + "" + P + "" + O
//Construct an equivalent Filter Constraint of the current criteria
          if (field_criteria[i-1].Type() = String) then
fieldConst += "?o" + i + "=" + " ' " + field_criteria[i-1] + " ' "
          else
          fieldConst += "?o" + i + "=" + field_criteria[i-1]
          end if
```

```
        if (fieldConst.isEmpty() = False AND i<L) then
        fieldConst += "&&"
        end if
    end for
else if (field_criteria.isValue() = True AND field.isComplex() = False) then
    //The first Triple Pattern
    S = subject; P = field; O = "?x"
    TP += S + "" + P + "" + O
    //The second Triple Pattern
    S=O; P= "rdf:member"; O= "?o"
    TP += S + "" + P + "" + O
    // Construct Filter Constraint
    if (field_criteria.type() = String) then
        fieldConst = "?o = ' " + field_criteria + " ' "
    else
        fieldConst = "?o =" + field_criteria
    end if
else if (field_criteria.isValue() = True AND field.isComplex() = True) then
    field_label = field.getLabel()
    field_index = field.getIndex()
    //The first Triple Pattern
    S=subject; P= field_label; O= "?x"
    TP += S + "" + P + "" + O
    //The second Triple Pattern
    S=O; P= "rdf:_"+ (field_index+1); O= "?o"
    TP += S + "" + P + "" + O
    // Construct Filter Constraint
    if (field_criteria.type() = String) then
        fieldConst = "?o = '" + field_criteria + "'"
    else
```

```
        fieldConst = "?o =" + field_criteria
    end if
end if
Awhere = Array("TP"==> TP,   "filter" ==> fieldConst) ;
return Awhere
End Algorithm
```

3.5 Embedded Document Field Converter:

This component adopts the same principle of the previous one (Array Field Converter). it takes as input an embedded document type field, its criteria and its relative subject. The processing of this algorithm is divided in two main parts checking two different conditions on the field criteria type if it is in the form of a simple value or an embedded document so as to return at the end an associative array of SPARQL Where Clause elements.

```
Input: subject; an embedded document type field, field; Criteria of this field, field_criteria
Output: An associative array of SPARQL Where Clause elements of the input field, EDwhere[]
Begin
TP, S, P, O, fieldConst, f, fc = ""; EDwhere[]; N=0;
if (field_criteria.isValue() = True) then
    field_label = field.getLabel()
    field_EDfield = field.getEDfield ()
    //The first Triple Pattern
    S=subject; P= field_label; O= "?x"
    TP += S + "" + P + "" + O
    //Other Triple Patterns
    S=O; P= field_EDfield; O= "?"+field_EDfield
    TP += S + "" + P + "" + O
     // Construct Filter Constraint
     if (field_criteria.type() = String) then
          fieldConst = O +"='" + field_criteria + "'"
     else
          fieldConst = O + "=" + field_criteria
     end if
else if (field_criteria.type() = EmbeddedDocument) then
```

```
   //The first Triple Pattern
   S=subject; P= field; O= "?x"
   TP += S + "" + P + "" + O
   N = field_criteria.countKVpairNbr();
   for (i=0; i<N; i++)
         f = field_criteria.getEDfield(i);
         fc = field_criteria.getEDfieldCriteria(i);
         S="?x"; P= f; O= "?"+P
       TP += S + "" + P + "" + O
         if (fc.type() = String) then
             fieldConst += O +"='" + field_criteria + "'"
         else
             fieldConst = O + "=" + field_criteria
         end if
   end for
   if (fc.isEmpty() = False AND i<N-1) then
         fc += "&&"
  end if
end if
EDwhere = Array("TP"==> TP,   "filter" ==> fieldConst) ;
Return EDwhere
End Algorithm
```

3.6 Construct SPARQL ORDER BY clause:

it takes as input a MongoDB query sort criteria in the form of Field : Order aims to specify ascending (if Order variable equals to 1) or descending (if Order variable equals to 1) sort on existing field so as to return at the end of this algorithm a SPARQL ORDER BY clause. In fact, it extracts the sort criteria parameters and verifies the order value; if it is equal to 1 then the SPARQL ORDER BY clause is written as ORDER BY ASC(?field), else if it is equal to -1 then the SPARQL ORDER BY clause is written as ORDER BY DESC(?field).

```
Input: A MongoDB query sort criteria, S
Output: A SPARQL ORDER BY Clause, orderBy
Begin
orderBy = "ORDER BY ";
field = S.getField();
order = S.getOrderValue();
if (order = 1) then
        orderBy += "ASC" + "(" + "?" + field + ")";
else if (order = -1) then
        orderBy += "DESC" + "(" + "?"  + field + ")";
end if
return orderBy
End Algorithm
```

3.7 Query Mapping:

The QueryMapping is the main procedure of our algorithm; it takes as input a MongoDB query so as to return at the end of treatments the SPARQL equivalent query. Firstly, this procedure starts with an initialization of output query elements (SELECT and WHERE clauses) on one hand, and the generation of the input query conversion tree by using the parse function. Regardless of the input query type, the SPARQL SELECT clause equal to "*" if the projection of input query is empty else its value is generated by the ConstructSparqlSELECTclause sub-procedure, and the SPARQL WHERE clause is constructed by the sub-procedure ConstructSparqlWHEREclause. If the input query type is "basic_query" then the output query is obtained by concatenating the SELECT and WHERE clause, else if the type is "sort_query" then we obtain the output query by concatenating the SPARQL SELECT and WHERE clause in addition to the keyword "ORDER BY" and its parameter generated by ConstractSparqlORDERBYclause sub-procedure. Likewise, if the type is "limit_query" then the output equivalent query is composed by SELECT and WHERE clause concatenating with the keyword "LIMIT" and its parameter obtained from the parse tree. In the case where the type is "mix_query", the output query will composed of SELECT and WHERE clause concatenated with ORDER BY and LIMIT clause.

Input: A MongoDB query, q_{in}

Output: A SPARQL query, q_{out}

Begin

q_{out} = "" {A SPARQL query that is initially blank}

tree = parse(q_{in})

$q_{in}^{criteria}$ = tree.getCriteria();

$q_{in}^{projection}$ = tree.getProjection();

q_{in}^{sort} = tree.getSortCriteria();

q_{in}^{limit} = tree.getLimitParam();

q_{out}^{SELECT} = "SELECT "; q_{out}^{WHERE} = "WHERE {"

if ($q_{in}^{projection}$.isEmpty() =true) **then**

q_{out}^{SELECT} += "*";

else

q_{out}^{SELECT}+= ConstructSparqlSELECTclause($q_{in}^{projection}$);

end if

q_{out}^{WHERE} += ConstructSparqlWHEREclause($q_{in}^{criteria}$) + "}";

if (q_{in}.type() = basic_query) **then**

q_{out} = select + where;

else if (q_{in}.type() = sort_query) **then**

q_{out} = select + where + ConstractSparqlORDERBYclause(q_{in}^{sort});

else if (q_{in}.type() = limit_query) **then**

q_{out} = select + where + "LIMIT " + q_{in}^{limit} ;

else if (q_{in}.type() = mix_query) **then**

q_{out} = select + where + ConstractSPARQLORDERBYClause(q_{in}^{sort}) +"LIMIT "+ q_{in}^{limit} ;

end if

return q_{out}

End Algorithm

4 Implementation

In order to implement our proposed mapping method, we have developed a tool named *Mongo2SPARQL* (for transforming **Mongo**DB query language to **SPARQL**) using Java language in order to be portable and

easy to evolve. Our tool is characterized by its simplicity and efficiency to allowing MongoDB users to integrate and query semantic data with MongoDB query language.

We present below some examples of MQL queries supported by our system *Mongo2SPARQL* (operate on the collection illustrated in Fig.2) and its equivalent SPARQL queries (operate on the equivalent RDF Schema designed in Fig.3).

- Example 1: presents a simple MQL find query (composed of three criteria, one projection parameter and sort method) and its equivalent SPARQL query (Fig.4).
- Example 2: treats an MQL query aiming to match an Array element (the first element of an array field named experiences) in order to convert it to a SPARQL equivalent query as illustrated in Fig.5.
- Example 3: illustrates the conversion of an MQL query with an exact matching on an Embedded Document, to a SPARQL equivalent query as shown in Fig.6.

```
{ _id : “http:://fsts.com/phdStudent /100”
name : “Abdeljalil Boumlik”,
age : 26,
status: “phd student”,
university : “FST”,
experiences : [ “Senior Programmer”, “Support Engineer”, “System Analyst” ],
diplomas : [   {diploma : “Bac” , mention : “pretty good”},
                {diploma : “Licence” , mention : “good”},
                    {diploma : “Master” , mention : “good”} ]
}
{ _id : “http:://fsts.com/phdStudent/101”,
name : “Nassima Soussi”,
age : 24,
status: “phd student”,
university : “FST”,
experiences : [ “DBA”, “Computer Engineer”, “Web Designer”],
diplomas : [ { diploma : “Bac” , mention : “good” } ,
                { diploma : “Engineering”, mention : “good” } ]
}
```

Figure 2. Doctorate Collection

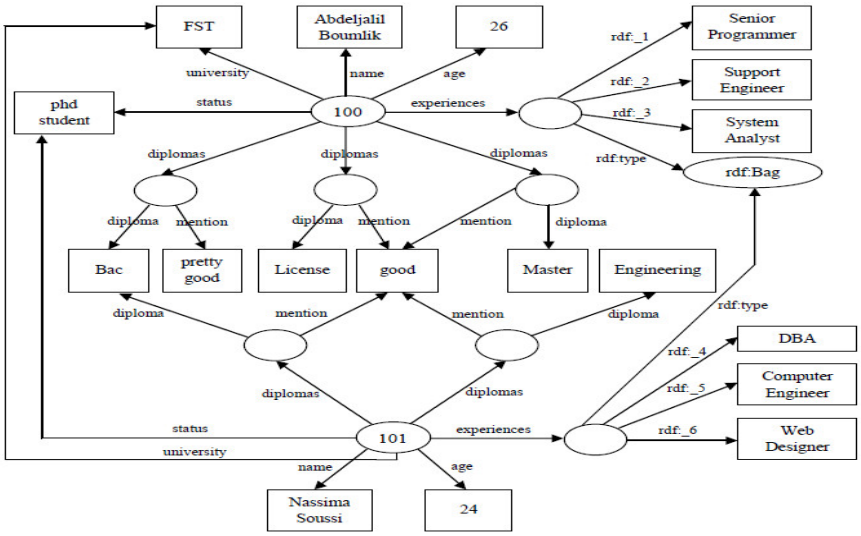


Figure 3.RDF Schema

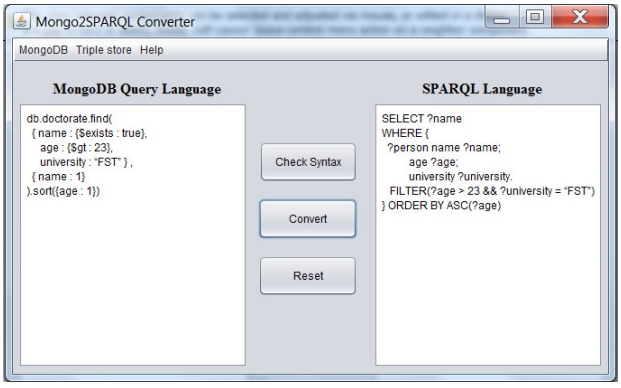


Figure 4. Example of Simple MQL find Query

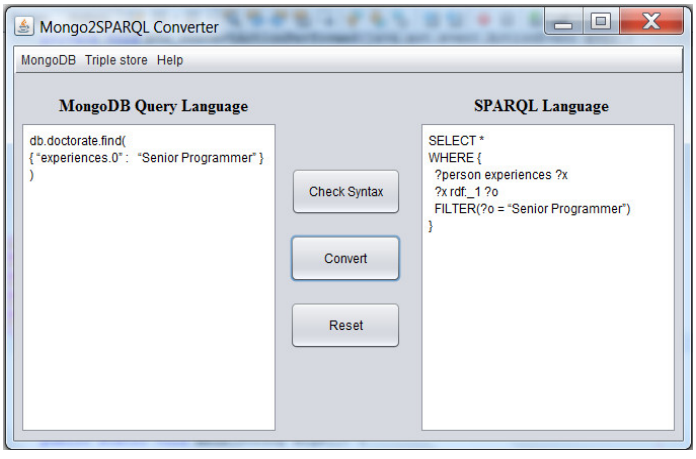


Figure 5. Example of Matching an Array Element in MQL Query

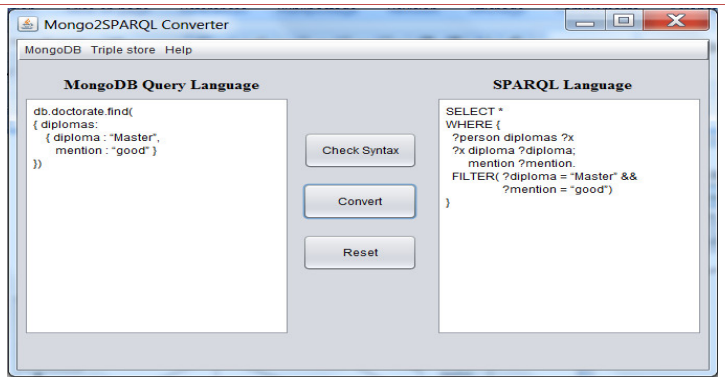


Figure 6.Example of an Exact Matching of an Embedded Document in MQL Query

5 Conclusion

In order to bridge the gap between NoSQL world and Semantic web world, we have established in this paper the first approach allowing to look at RDF stores trough MongoDB lenses by elaborating an efficient conversion query algorithm named Mongo2SPARQL that transform each component of find, sort and limit MongoDB query methods to its equivalent components in SPARQL language taking into a count the different types of MQL Criteria Field (simple value, array, embedded document). We have also implemented our algorithm with a java language.

One obvious extension of our research is to enhance our algorithm so as to support more MongoDB collection query types of *find* method such as *findAndModify*, *findOne*, *findOneAndDelete*, *findOneAndReplace* and *findOneAndUpdate*. Another promise about our future works is to integrate this conversion technique to MongoDB database system aiming to offer NoSQL users, more precisely, MongoDB users an easy bridge and an open extension toward semantic world.

REFERENCES

[1] Tudorica, B. G., & Bucur, C. (2011, June). A comparison between several NoSQL databases with comments and notes. In Roedunet International Conference (RoEduNet), 2011 10th (pp. 1-5). IEEE.

[2] Prud'hommeaux, E., & Seaborne, A. (2008). SPARQL query language for RDF. Recommendation, W3C. Retrieved from https://www.w3.org/TR/rdf-sparql-query/. Last visit February 2016.

[3] The MongoDB 3.2 Manual. Retrieved from https://docs.MongoDB.org/manual. Last visit February 2016.

[4] Tomaszuk, D. (2010). Document-oriented triple store based on RDF/JSON.Studies in Logic, Grammar and Rhetoric,(22 (35)).

[5] Curé, O., Lamolle, M., & Duc, C. L. (2013). Ontology based data integration over document and column family oriented NOSQL. arXiv preprint arXiv:1307.2603.

[6] Michel, F., Faron-Zucker, C., & Montagnat, J. (2016). Mapping-based SPARQL access to a MongoDB database (Doctoral dissertation, CNRS).

[7] Rocha, L., Vale, F., Cirilo, E., Barbosa, D., & Mourão, F. (2015). A Framework for Migrating Relational Datasets to NoSQL. Procedia Computer Science, 51, 2593-2602.

[8] Lawrence, R. (2014, March). Integration and virtualization of relational SQL and NoSQL systems including MySQL and MongoDB. In Computational Science and Computational Intelligence (CSCI), 2014 International Conference on (Vol. 1, pp. 285-290). IEEE.

[9] Khan, S., & Mane, V. (2013). SQL support over MongoDB using metadata.International Journal of Scientific and Research Publications, 3(10), 1-5.

[10] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 28-37.

[11] Beckett, D. (2014, February). Rdf 1.1 xml syntax. W3C recommendation. Retrieved from https://www.w3.org/TR/rdf-syntax-grammar/. Last visit Fabruary 2016.

[12] Bechhofer, S., Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L.A.: OWL Web Ontology Language Reference. W3C Recommendation 10, 2006–10 (2004).

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Content-based Medical Image Retrieval for Liver CT Annotation

[1]Imane Nedjar, [2]Saïd Mahmoudi, [3]Mohammed Amine Chikh
[1,3]Biomedical Engineering Laboratory, Tlemcen University, Tlemcen, Algeria
[2]Computer Science Department, Faculty of Engineering, University of Mons, Mons, Belgium
imane.nedjar@mail.univ-tlemcen.dz, said.mahmoudi@umons.ac.be, ma_chikh@mail.univ-tlemcen.dz

ABSTRACT

The increase number of medical image stored and saved every day presents a unique opportunity for content-based medical image retrieval (CBMIR) systems. In this paper, we propose content-based medical image retrieval for annotating liver CT scans images in order to generate a structured report. For that, we have used the Bidimentional Empirical Mode Decomposition (BEMD), and then we have applied Gabor wavelet transform to extract the mean and the standard deviation as features descriptors. Finally, a proposed similarity distance was employed to retrieve the most similar training images to the image query, and a majority voting scheme was used to select the answers for an unannotated image. We have used the IMAGECLEF 2015 annotation dataset and the obtained score was 88.9%.

Keywords: Medical image, Liver annotation, Image Retrieval, BEMD, Gabor wavelet.

1 Introduction

The huge amount of digital images produced in hospitals and health care centers needs the development of automatic mechanisms for their storage and retrieval from databases. The structured reports are highly valuable in medical contexts due to the processing opportunities that they provide. However, structured reports are time consuming, since they need a lot of time to be created. Furthermore, their creation requires high domain expertise. Consequently, such structured medical reports are often not found or are incomplete in practice [1].

The goal of this work is to improve computer-aided automatic annotation of liver CT volumes by filling in a structured radiology report. Therefore, the major part of this work is devoted to describe the semantic features of the liver, its vascularity, and the types of lesions in the liver.

2 Related works

Several schemes and algorithms have been presented in the literature for annotation of liver CT scans.

In 2015 the authors of [2], proposed two methods for annotating the liver image. The first one uses a classification approach composed of two main phases. The first step consists of a pre-processing allowing the extraction of texture and shape based features vector from both liver and lesion. Afterward, a classification process is achieved by using random forest classifier. The second method uses a specific signature of the liver. This method is applied to a set of slices from 3D liver CT scans which are normalized into rectangular blocks. After applying the 1D Log-Gabor filters transformation dominant phase data of each block is extracted and quantized to four levels in order to encode the unique pattern of the liver into

DOI: 10.14738/tmlai.54.2985
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2985

a bit-wise template. The Hamming distance is employed for retrieval. The proposed methods were evaluated on ImageCLEF2015 liver CT annotation dataset, all these methods achieved scores (>90), and a completeness level of 99%.The best score was 91% given by the second method.

In 2014, Kumar et al [3] proposed two strategies for annotating the liver images. The first method uses multi-class classification scheme where each label has a classifier trained to separate it from other labels. The second method uses the similarity scores from an image retrieval algorithm as weights for a majority voting scheme, thereby, reducing the inherent bias towards labels that have a high number of samples. The proposed methods were evaluated by using ImageCLEF2014 liver CT annotation dataset, they achieved the highest score of 94.7% out of all the submission done to imageCLEF2014.

On the other side, the authors in [4] tried four different classifiers in the learning phase: linear discriminant analysis (LDA), logistic regression (LR), K-nearest neighbors (KNN), and finally SVM, to predict labels. An exhaustive search of every combination of image features is done by using leave-one-out cross validation method on training data for every label and classifier. Python scikit-Learn Machine learning toolbox was used for implementing each classifier with default parameters. As result, for most of the labels they got almost the same performance by using any classifier and any combination of image features. The proposed method was evaluated with ImageCLEF2014 liver CT annotation dataset, and they obtained a score of 93%.

In [5] the authors proposed an approach based of a probabilistic interpretation of tensor factorization models, i.e. Generalized Coupled Tensor Factorization (GCTF). This method can simultaneously fit a large class of matrix/tensor models to higher-order matrices/tensors with common latent factors using different loss functions. The authors considered the dataset as heterogeneous data, and the GCTF approach was applied to predict labels. They considered KLdivergence and Euclidean-distance based cost functions as well as the coupled matrix factorization models by using the GCTF framework. The proposed method was evaluated with ImageCLEF2014 liver CT annotation dataset, and their highest score was 67.7%.

3 Material

In our experimentation, we have used the ImageCLEF2015 liver CT annotation dataset. The training dataset includes 50 cases, each consisting of:

- A cropped CT image of the liver, a 3D matrix. The volumes had varied resolutions (x: 190-308 pixels, y: 213-387 pixels, slices: 41-588) and spacing (x, y: 0.674-1.007mm, slice: 0.399-2.5mm),
- A liver mask that specifies the part corresponding to the liver, A bounding box (ROI) corresponding to the selected region of the lesion within the liver,
- An RDF file generated using Liver Case Ontology (LiCO) representing manually entered imaging observations by a radiologist. In total, there are 73 User Expressed Features (UsE). These features clinically characterize the liver, hepatic vascularity, and liver's lesions, an example of UsE is shown in the following table.

Table 1. Exemple of UsE features [6]

Group	Concept	Properties	Possible values (assigned indices)
Liver	Liver	Liver Placement	downward displacement(0), normal placement(1), leftward displacement(2), upward displacement(3), other(4)

The test dataset contained 10 CT volumes, with varied resolutions and pixel spacing, cropped to the region around the liver. The test data also included a mask of the liver pixels and a bounding box (ROI) (see Fig. 1). But, there is no RDF file and imaging observations are missing.

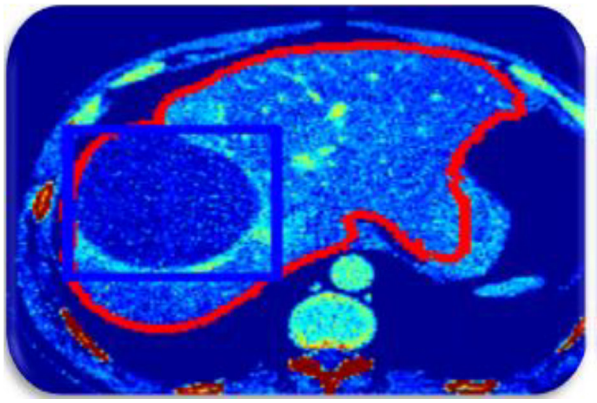


Figure 1. Image test of the liver, the contour in red and ROI in blue

4 The proposed method

Our goal in this work was to investigate a new annotation process using content based image retrieval. The proposed approach is based on a spectral images textural analysis that uses a combination of Bidimensional Empirical Mode Decomposition (BEMD) and Gabor wavelet transform. The features vector includes the mean and the standard deviation of each filtered image.

Thereafter, based on a similarity distance we selected the five most similar images, and the label that has the majority voting will be assigned to the User Expressed Features (UsE).

We used the proposed similarity score calculation between the unannotated image (U) and a training image (T) as the distance between their respective features vectors is given as bellow in equation (1):

$$Sim(U,T) = \sum_{i=0}^{d} 1 - \frac{|u_i - t_i|}{v_i} \quad (1)$$

Where v_i is the i-th maximum feature in dataset, u_i is the i-th feature in the feature vector of U, t_i is the i-th feature in the feature vector of T, and d presents the dimensionality of the feature set.

In the following section, we present a general overview of the different methods used in this paper.

4.1 Bidimensional Empirical Mode Decomposition (BEMD)

The Bidimensional Empirical Mode Decomposition (BEMD) approach is a highly adaptive decomposition [7]. It is mainly based on the characterization of image by Intrinsic Mode Function (IMF) decomposition, where the image can be decomposed into a redundant set of composite images called IMF and a residue. Adding all the IMFs with the residue allow reconstructing the original image without distortion or loss of information [8]. This method, derived from image data and which is fully unsupervised, permits to analyze nonlinear and non-stationary data as texture images.

4.2 Gabor Wavelet Transform

The Gabor wavelet representation allows good recognition without correspondence, because it captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity [9]; it is defined as follows in equation (2):

$$\psi_{u,v}(z) = \frac{\|K_{u,v}\|^2}{\sigma^2} e^{\left(\|K_{u,v}\|^2 |z|^2 / 2\sigma^2\right)} \left[e^{ik_{u,v}z} - e^{-\sigma^2/2}\right] \quad (2)$$

Where u and v define the orientation and the scale of the Gabor kernels, and ||.|| denotes the norm operator, and the wave vector K_{uv} is defined as follows:

$$K_{u,v} = K_v e^{i\theta\mu} \quad (3)$$

Where $K_v = Kmax/f_v$ and $\theta_u = \pi u / 8$, Z=(x,y) and Kmax is the maximum frequency, f is the spacing factor between kernels in the frequency domain. We have used $\sigma = 2\pi$, $K_{max} = \pi/2$, and $f = 2^{1/2}$.

Let I(x,y) be the gray level distribution of an image, the convolution of the image I and a Gabor kernel $\psi_{u,v}$ is defined as follows:

$$O_{u,v} = I(z) * \psi_{u,v}(z) \quad (4)$$

Where * denotes the convolution operator, and $O_{u,v}(z)$ is the convolution result corresponding to the Gabor kernel at orientation u and scale v.

The energy information of the image I (x, y) can be expressed as follows[10]:

$$E(u,v) = \sum_x \sum_y \left|O_{u,v}\right| \quad (5)$$

In this work, we propose to use the mean and the standard deviation as texture features. Suppose that I(x,y) denotes an image of M × N pixels, $\mu_{uv}(x)$ and $\sigma_{uv}(x)$ denote its mean and standard deviation computed from the scale v and the direction u, respectively, $\mu_{uv}(x)$ and $\sigma_{uv}(x)$ can be computed as follows:

$$\mu_{u,v} = E_{u,v} / MN \quad (6)$$

$$\sigma_{u,v} = \sqrt{\frac{\sum_x \sum_y (\left|O_{u,v}(z) - \mu_{u,v}\right|)^2}{MN}} \quad (7)$$

5 The proposed method

In order to evaluate our approach, we calculated two descriptors:

Descriptor n°1: computed by applying Gabor wavelet transform with sixteen orientations and ten scales we have extracted 160 elements of means and 160 elements of deviation from the first BIMF. Therefore, the total dimension of the first descriptor is 320.

Descriptor n°2: calculated from the first, the second and the third BIMF, we have extracted 160 elements of means and 160 elements of deviation for each one of them. Therefore, the total dimension of the second descriptor is 960.

5.1 Evaluation methodology

The evaluation of our experiments is performed on the basis of completeness and accuracy of the predicted annotations, with reference to the manual annotations of the test dataset. Completeness is defined as the number of predicted features divided by the total number of features, equation (8), while accuracy is the number of correctly predicted features divided by the total number of predicted features, equation (9).

$$\text{Completeness} = \frac{\text{number of predicted UsE features}}{\text{Total number of UsE features}} \tag{8}$$

$$\text{Accuracy} = \frac{\text{number of corrected predicted UsE features}}{\text{Number of predicted UsE features}} \tag{9}$$

$$\text{TotalScore} = \sqrt{\text{Completeness} \times \text{Accuracy}} \tag{10}$$

The following table shows the score results obtained by the two proposed descriptors applied for predicting different groups of UsE features.

Table 2.The score results obtained for each group

	Descriptor n°1		**Descriptor n°2**	
	Cmp	*Acc*	*Cmp*	*Acc*
Liver	1.00	0.92	1.00	0.92
Vessel	1.00	1.00	1.00	1.00
Lesion Area	1.00	0.65	1.00	0.67
Lesion Lesion	1.00	0.48	1.00	0.51
Lesion component	0.96	0.86	0.96	0.89
Total score	**0.880**		**0.889**	

The results presented in table 2 show that the scores obtained by the two descriptors for Liver group and Vessel group are the same. This can be explained by the fact that there were some instances having the same annotation (unbalanced dataset) in all the training samples, and this is the major challenge in this dataset. We notice that the second descriptor outperforms the first descriptor in the other groups. The

best result is obtained by using three BIMFS with the score of 88.9% compared to the first descriptor with the score of 88%. This suggests that, using more details improves the accuracy of the annotation.

Table 3 shows that all authors have completed the vessel UsE features with high accuracy, also they completed the liver UsE features with accuracy more than 80% except the method proposed in [5].

Concepts related to lesion-lesion are annotated completely by[2] [3] and our proposed approach. However, [3] has got the accuracy of 83% by using the second feature set of size1446, while we achieved 51%.We attribute this difference to the size of feature set which has significantly improved the accuracy of the annotation.

Lesion-components group are fully completed and annotated with an accuracy higher than 90% by both [3] and [4]. In our proposed approach we achieved a completeness score of 96%, and accuracy of 89%.

It is interesting to note that one of the major considerations in the annotation of the liver is the appropriate combination of features and methods.

Table 3.Comparisontest results presented by different authors and summarized in [6]

Methods presented In	Liver		Vessel		Lesion-Area		Lesion-Lesion		Lesion-Component	
	Cmp	*Acc*	*Cmp*	*Acc*	*Cmp*	*Acc*	*Cmp*	*Acc*	*Cmp*	*Acc*
[2]	1.00	0.92	1.00	1.00	1.00	0.75	1.00	0.48	0.96	0.89
[3]	1.00	0.93	1.00	1.00	0.92	0.79	1.00	0.83	1.00	0.94
[4]	1.00	0.93	1.00	1.00	0.85	0.81	0.90	0.82	1.00	0.94
[5]	0.62	0.88	1.00	1.00	0.46	0.77	0.20	1.00	0.12	0.15
Descriptor 2	1.00	0.92	1.00	1.00	1.00	0.67	1.00	0.51	0.96	0.89

6 Conclusion

In this paper, we have presented a new approach aiming to annotate liver CT scans with the use of content-based medical image retrieval. We first extracted a set of features descriptors based on spectral images textural analysis that uses a combination of Bidimensional Empirical Mode Decomposition (BEMD) and Gabor wavelet transform, finally the five most similar training images was used to select the answers for an unannotated image. Our experiments results have been conducted by using the ImageClef 2015 annotation dataset. Our best score of 88.9% was achieved by the second descriptor. In our futures works we plan to use more image features and examining others classifiers.

ACKNOWLEDGMENT

The authors would like to thank Mrs Neda Marvasti from Bogaziçi University, Electrical and Electronics Department, Istanbul, Turkey , for making the database available for the experiments.

REFERENCES

[1] N.Marvasti et al., "ImageCLEF Liver CT Image Annotation Task 2014. In: CLEF 2014 Evaluation Labs and Workshop", Online Working Notes. (2014).

[2] I.Nedjar, S.Mahmoudi, A.Chikh, K.Abi-ayad, and Z.Bouafia, "Automatic annotation of liver CT image:ImageCLEFmed 2015," in CLEF2015 Working Note.CEUR Workshop Proceedings,CEUR-WS.org, Toulouse,france, Septembre 8-11 2015.

[3] A.Kumar, S.Dyer, C.Li, P.H.W.Leong, and J.Kim, "Automatic annotation of liver ct images: the submission of the bmet group to imageclefmed 2014", in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR, Workshop Proceedings (CEUR-WS.org), September 2014.

[4] A.B.Spanier and L.Joskowicz, "Towards content-based image retrieval: From computer generated features to semantic descriptions of liver ct scans", in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

[5] B.Ermis and A.T.Cemgil, "Liver ct annotation via generalized coupled tensor factorization", in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

[6] N.Marvasti et al., "Overview of the ImageCLEF 2015 liver CT annotation task," in CLEF2015 Working Notes. Workshop Proceedings.CEUR-WS.org, no. 1613-0073, September 8-11. 2015.

[7] N.E.Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. London. A 454, 903-995, 1998.

[8] J.C.Nunes, Y.Bouaoune, E.Delechelle, O.Niang, and P.Bunel, "Image analysis by bidimensional empirical mode decomposition," Image and Vision Computing 21 1019–1026, 2003.

[9] C.Liu and H.Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," IEEE Transactions on Image Processing, vol. 11, no. 4, pp. 467-476, 2002.

[10] G.Zhang, ZM.Ma, and L.Deng, "Texture Feature Extraction and Description Using Fuzzy Set of Main Dominant Directions of Variable Scales in Content-Based Medical Image Retrieval," in SAC '08 Proceedings of the ACM symposium on Applied computing, Fortaleza, Ceará, Brazil, March 16-20, 2008, pp. 1760-1761.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

A Model Driven Architecture Approach to Generate Multidimensional Schemas of Data Warehouses

[1]O. Betari, [1]M. Erramdani, [2]K. Arrhioui
[1]MATSI Laboratory, Superior School of Technology, Mohammed First University, Oujda, Morocco
[2]MISC Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco
beta.oualid@gmail.com, m.erramdani@gmail.com, arr.karim@gmail.com

ABSTRACT

Over the past decade, the concept of data warehousing has been widely accepted. The main reason for building data warehouses is to improve the quality of information in order to achieve specific business objectives such as competitive advantage or improved decision-making. However, there is no formal method for deriving a multidimensional schema from heterogeneous databases that is recognized as a standard by the OMG and the professionals of the field. Which is why, in this paper, we present a model-driven approach (MDA) for the design of data warehouses. To apply the MDA approach to the Data warehouse construction process, we describe a multidimensional meta-model and specify a set of transformations from a UML meta-model which is mapped to a multidimensional meta-model. The execution of the transformation, programmed by the Query View Transformation (QVT) language, takes as input an instance of the UML Meta-Model to generate an instance of the Dimensional Meta-Model as output.

Keywords-Data Warehouse; Meta model; Model Driven Architecture; Transformation.

1 Introduction

Decision-making systems are defined by information systems that can help companies in their decision-making process. To support these decisions, it has become necessary to manage and analyze large amounts of data to obtain new and relevant knowledge. In this sense, a new database concept was created: Data Warehouse [1]. However, building a DW is still a challenging and complex task because it consists of several interrelated components with different functions, different design pitfalls, and different technologies [2].

To overcome this problem, we decided to use the Model Driven Architecture (MDA), an approach developed by Object Management Group (OMG), in building the DW. MDA addresses the challenges of constantly evolving highly interlinked systems, providing an architecture that provides portability, interoperability across platforms, platform independence, domain specificity, and productivity [3]. This framework separates the specification of system functionality in a Platform Independent Model (PIM) from the specification of the implementation of that functionality on a specific technology in a Platform Specific Model (PSM). Furthermore, the system requirements are specified in a Computation Independent

DOI: 10.14738/tmlai.54.2397
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.2397

Model (CIM). The main benefit of MDA is that less time and effort are needed to develop the whole software system, thus improving productivity [4]-[7].

Considering these aspects, the present paper describes a PIM to PIM transformation using an approach by modeling, we will automatically generate DW schema by using several UML profiles. The resulting PIM represent conceptual model of DW component, without any reference to a concrete target platform or a specific technology. We developed as well the transformation rules by using the formal QVT transformations.

The present paper is structured as follows. Section 2 presents the most relevant related works. In section 3 we outline the MDE principles. Section 4 explains the concepts of Data warehousing. Sections 5 introduces our MDA approach for generating the MD modeling of the DW repository. Finally, section 6 concludes the work and offers further perspectives.

2 Related Works

In this section, we present a brief overview of some approaches, proposed for the development of DW systems.

In [8], the authors propose a Fact-Dimension model. It is a graphical conceptual model for DW. They also show how to derive a DWH schema from the data sources described by the entity-relationship diagram.

A multidimensional conceptual Object-Oriented model for Data Warehousing is proposed in [9]. It was developed as an extension of the UML notation to represent the multi-dimensional data structure. The authors also present OLAP (OnLine Analytical Processing) tools, its structures, integrity constraints and query operation but they don't show how to get the presented conceptual model.

In [1], the authors are interested in secure XML data warehouses. They propose an approach for the model driven development of these DW in which the secure conceptual DW data model, the PIM, is transformed into a secure XML DW, as a PSM, by applying a set of transformation rules.

In [4], the authors present an MDA approach for the development of the DW repository, and then, describe how to build the different MDA models for the DW repository by using an extension of UML and the Common Warehouse Meta model (CWM). They modeled the PIM by using their UML profile for the multidimensional modeling of DW. They derive then the PSM taking into account the deployment platform. From this PSM, the necessary SQL code is derived to create data structures for the DW in a relational platform.

In [10], a well-structured approach is proposed to formalize the development of the DW repository. Authors establish a set of multidimensional normal forms in order to obtain a conceptual model of the DW repository. A relational representation from the conceptual model is informally derived, but no formal transformations are defined to obtain the logical model.

3 Model Driven Engineering

3.1 The OMG approach

In late 2000, OMG, a consortium of over 1,000 companies, first reviewed the document entitled "Model Driven Architecture" and decided to form an architecture team to produce a more formal statement of

the MDA [3]. This approach focuses on developing the highest level of abstraction models and promotes the transformation approach from one model to another.

MDA addresses the challenges of today's highly networked, constantly changing systems, providing an architecture that assures portability, cross-platform Interoperability, platform independence, domain specificity and productivity [3]. The key to the MDA approach is the importance of models in the software development process. In the MDA, the software development process is driven by the modeling activity of the software system.

In the field of software engineering, the OMG with its MDA approach classifies four types of models that it advocates for the construction of software: Computation Independent Model (CIM), Platform Independent Model (PIM), Platform Specific Model (PSM) and Code; defined as follow:

- CIM: The goal is to create a requirements model for the future application. Such a model must represent the application in its environment in order to define the services offered by the application and which other entities with which it interacts.
- PIM: This model represents the system-specific business logic or the design model. It represents the functioning of entities and services. It must be durable and lasting over time. It describes the system, but does not show the details of its use on the platform.

PSM: MDA considers that the code of an application can be easily obtained from these models. The main difference between a code model and an analysis or design model is that the code model is linked to an execution platform.

The reason for the above model organization is to develop models of the systems' business logic independently from the platforms of execution, then to transform these models automatically to models dependent of the platforms. The complexity of the platforms does no longer appear in the business logic models but it' is found in the transformation [11].

The required steps during the model-driven development with the UML approach can basically be divided into the following steps [12], at first building the CIM that acquires user requirements. Then, according to this CIM, a PIM is built. Next is, the transformation of the proposed PIM into one or more PSMs. This type of transition from CIM to PIM and PIM to PSM is called Model To Model (M2M) transformation. The final step is to transform the generated model respecting the PSM into the code of the chosen platform. This transition is called Model To Text (M2T) transformation. Figure 1, shows how the transformations are done.

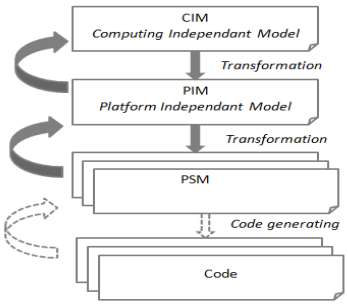


Figure 1. Model Driven Architecture levels

The OMG, in the context of MDA, gives the following definition to the transformation of models "the process of converting a model into another model of the same system". In general, it can be said that a transformation definition consists of a collection of transformation rules, which are unambiguous specifications of how a model can be used to create another model. There are three approaches in MDA to perform these transformations:

- Approach by programming: Consists in using object-oriented programming languages. The idea is to program a transformation of models in the same way that any computer application is programmed.
- Approach by template: Consists in defining the canvas of the desired target models by declaring parameters. These parameters will be substituted by the information contained in the source models. The execution of a transformation involves taking a template and replacing its parameters with the values of a source model.
- Approach by Modeling: Consists in applying the concepts of the engineering of the models to the transformations of the models themselves. The objective is to model the transformations of models and to make the transformation models perennial and productive and to express their independence vis-à-vis the execution platforms.

Using the modeling approach is designed to have a sustainable and productive models' transformation, independently of any execution platform. This is why the OMG has developed a standard for this transformation language which is the MOF 2.0 QVT [14], standing for Query View Transformation.

4 Data Warehouses

In the early 1990s, as a consequence of a competitive world, organizations realized that data analysis needs to be performed to support their decision making processes. Traditional transactional databases do not satisfy the requirements for data analysis because they are designed to support daily operations and ensure concurrent access. However, they do not include historical data and are not optimized to execute complex queries that involve large volumes of data. Therefore, data warehouses were proposed as a solution to overcome these problems and limitations [15]. A data warehouse provides an infrastructure that allows users to get efficient and accurate responses to complex queries. Different systems and tools can be used to access and analyze data stored in data warehouses.

A data warehouse is a subject-oriented, integrated, non-volatile, and time-variant collection of data in support of management's decisions [16]:

- Subject-oriented: The Data Warehouse is organized around major topics of the company such as: customer, sales, products… This organization necessarily affects the design and implementation of data in the Data Warehouse.
- Integrated: Data is integrated from various operational and external systems.
- Non-volatile: Data is accumulated from operational systems for a long period of time.
- Time-variant: In a decision-making system it is important to keep the different values of a given data, allowing comparisons and monitoring of the evolution of values over time, the Data Warehouse keeps track of how its data evolved over time.

Figure 2, describes a typical architecture of a DW system, which consist of several tiers [17]:

- The back-end tier is composed of extraction-transformation-loading (ETL) tools, used to feed data in from operational databases and other data sources, and a data staging area, which is an intermediate database.
- The DW tier is composed of an enterprise data warehouse and/or several data marts, and a metadata repository.
- The OLAP tier is an OLAP server that supports multidimensional data and operations.
- The front-end tier contains client tools such as reporting tools and data-mining tools.

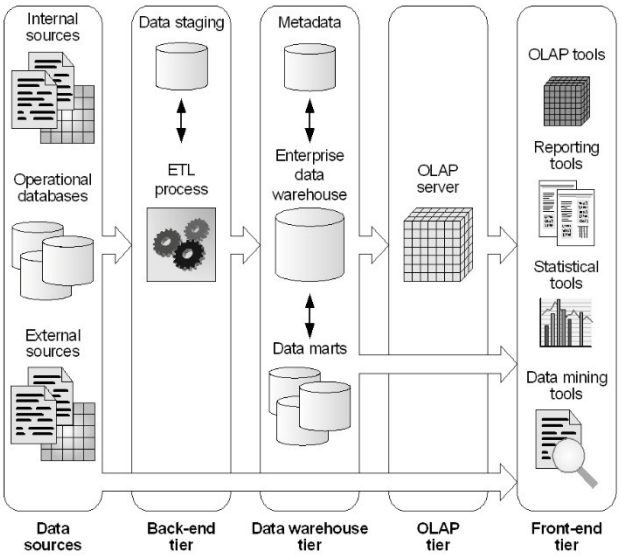


Figure 1. Typical data warehouse architecture

5 MDA for MD Modeling of DW

In this paper we proposed a meta model as a PIM to describe DW. Then, we chose to transform a UML model to PIM, with an approach by modeling using QVT. This type of transformation will allow us to automatically generate the multidimensional data warehouse schema from UML schema. We used UML in our approach, because it takes into account the structural and dynamic properties of the information system at a conceptual level. The class diagram of analysis is obtained from the data dictionary and functional dependencies. Obtaining the conceptual class diagram is based on scenarios of different use cases.

5.1 The Modeling process

5.1.1 PIM source meta model

The UML specification has the Kernel package which provides the core modeling concepts of the UML, our source Meta-model structures a simplified UML model based on a package containing the data types and classes. The classes contain structural features represented by attributes, and behavioral features represented by operations [13]. Fig. 3, presents the source meta model.

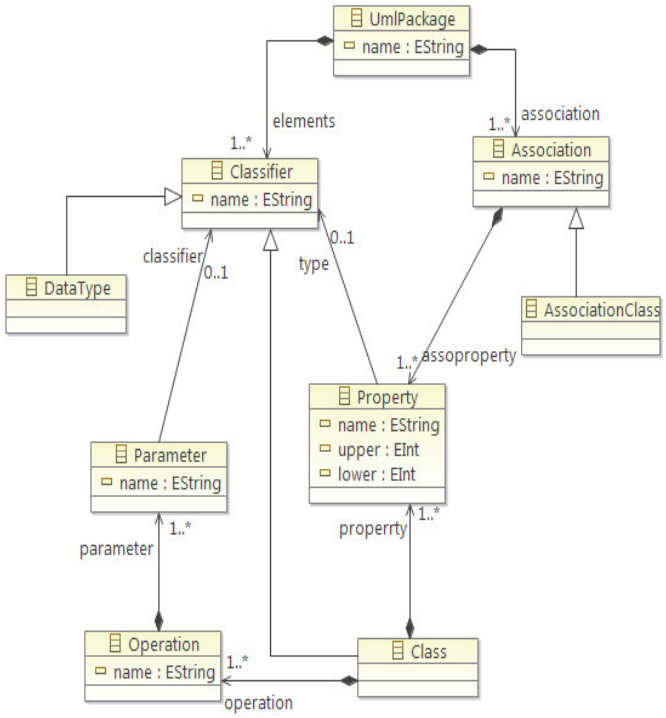


Figure 1. Simplified UML Meta-Model

- UmlPackage: is an abstract element of UML used to group together elements that are semantically related. This meta-class is connected to the meta-class Classifier.
- Classifier: this is an abstract meta-class which describes set of instances having common features. This meta-class represents both the concept of class and the concept of data type.
- Class: represents the concept of UML class.
- DataType: represents the data type of UML.
- Operation: expresses the concept of operations of a UML class.
- Parameter: describes the order, type and direction of arguments that can be given when the operation is invoked. It explains the link between Parameter meta-class and Classifier meta-class.
- Property: expresses the concept of Property of a UML class. These properties are represented by the multiplicities and meta-attributes upper and lower. A UML class is composed of properties, which explains the link between the meta-class Property and meta-class Class. These properties can be primitive type or class type. This explains the link between the meta-class and the Classifier meta-class Property.

5.1.2 PIM target meta model

Based on the approach described in [18], and inspired by this UML profile for the data warehouse, the target meta-model that we have proposed for the transformation, is a simplified meta-model containing Facts and Dimensions. A Fact contains attributes related to dimensions. This meta model is an extension of the work done by the authors in [19]. Fig. 4, shows the Target Meta Model:

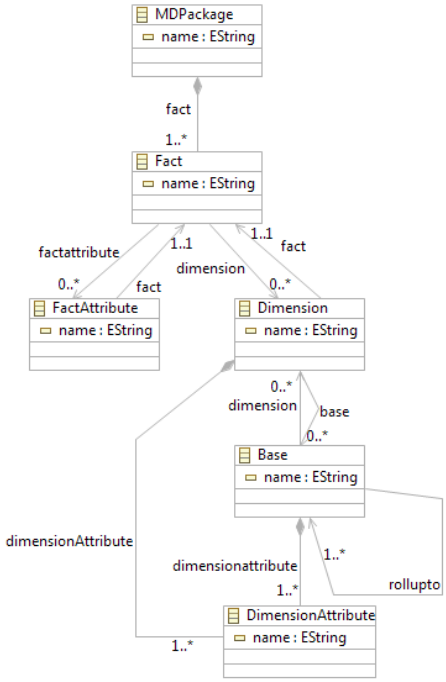


Figure 1. Simplified Meta-Model of multidimensional schema

- MDPackage: expresses the concept of the package, it contains Fact classes.
- Fact: expresses the concept of Fact in a multidimensional schema. Each fact is contained in MDPackage, which explains the link between MDPackage and Fact.
- FactAttribute: expresses the concept of attributes for the element Fact and Fact contains FactAttribute.
- Dimension: expresses the concept of dimension in a multidimensional schema.
- DimensionAttribute: expresses the concept of dimensions attributes. A dimension contains attributes, which explains the link between a Dimension Attribute and Dimension.
- Base: is the basic concept in the multidimensional schema. A base represents the hierarchical level of classification. An association (represented by the ROLLS-UPTO stereotype) between the base class specifies the relationship between two hierarchical levels of classification.

5.2 The transformation process

Once the meta models developed, the next step is to specify the correspondences between the two metamodels and automatically generateDW model from the UML model by using transformation rules written in QVT Operational Mappings. In this section we'll explain some of the programmed transformation rules.

A QVT transformation is in the form of a function with two parameters, the first corresponds to the source model, the second corresponds to the target model. The main function of our transformations is as follow:

```
transformationumlToDwhTransfo(insrc:UML, outdest:DWH);
main() {
      src.objectsOfType(UmlPackage) ->mapumlPackToDWHPack();
}
```

We consider that a simplified UML model consists of a package called UmlPackage. This package will be transformed into a multidimensional package called MDPackage, using the transformation rule defined as follows:

```
mappingUML::UmlPackage::umlPackToDWHPack() : DWH::MDPackage {
      result.name := self.name + "Dwh";
      result.fact += self.association[UML::AssociationClass] -
>mapassociationClassToFact();
}
```

The following code shows the transformation rule of a class association from a UML class diagram to a fact. The fact will carry the name of the association class, its attributes will be the properties of the association class:

```
mappingAssociationClass::associationClassToFact() : Fact {
      result.name := self.name + "Fact";
      factattribute += self.assoproperty ->mappropertyToFactAttribute();
}
```

6 Conclusion

In this paper, we have introduced our MDA approach for the development of DWs. For constructing the system, we applied transformations from a UML PIM in order to automatically obtain an analogous DW PIM. Thanks to the use of MDA and QVT, we can generate from a simplified UML model instance, a simplified Multidimensional model in just two steps: first is the development of the source and target PIMs; and second is the development of the corresponding QVT transformations.

The primary benefit of our approach is that the complex task of designing the whole DW is done in a systematic, well-structured, and standard way by using an MDA approach.

Our future intentions include implementing our MDA approach for the development of DWs into specific platforms by creating the corresponding PSMs. And transforming each generated PSM into code.

REFERENCES

[1] B. Vela, C. Blanco, E.F. Medina, E. Marcos, “A Practical Application of our MDD Approach for Modeling Secure XML Data Warehouses,” in Decision Support Systems (DSS), 54 (4), pp. 899-925, 2012.

[2] R. Kimball, M. Ross, The Data Warehouse Toolkit, 3rded., John Wiley & Sons, Inc. 2013.

[3] Object Management Group (OMG), MDA Guide 2.0. http://www.omg.org/cgi-bin/doc?ormsc/14-06-01

[4] J.N. Mazón, J. Trujillo, "An MDA Approach for the Development of Data Warehouses," in Decision Support Systems (DSS), 45 (1), pp. 41-58, 2008.

[5] D.S. Frankel, Model Driven Architecture: Applying MDA to Enterprise Computing, Wiley, 2003.

[6] A.G. Kleppe, J. Warmer, W. Bast, MDA Explained: The Model Driven Architecture: Practice and Promise, Addison-Wesley, 2003.

[7] S. Mellor, K. Scott, A. Uhl, D. Weise, MDA Distilled: Principles of Model-driven Architecture,Addison-Wesley, 2004.

[8] M. Golfarelli, D. Maio, S. Rizzi, "The Dimensional Fact Model: a Conceptual Model for Data Warehouses," in International Journal of Cooperative Information Systems (IJCIS), 7 (2&3), pp. 215-247, 1998.

[9] A. Abelló, J. Samos, F. Saltor, "YAM2: a Multidimensional Conceptual Model Extending UML," in Information Systems (IS), 31 (6), pp. 541-567, 2006.

[10] J. Lechtenbörger, G. Vossen, "Multidimensional Normal Forms for Data Warehouse Design," in Information Systems (IS), 25 (5), pp. 415-434, 2003.

[11] X. Blanc, MDA en Action: Ingénierie Logicielle Guidée par les Modèles, Eyrolles, 2005.

[12] S. Roubi, M. Erramdani, S. Mbarki, "Generating Graphical User Interfaces Based on Model Driven Engineering," in International Review on Computers and Software (IRECOS), 10 (5), pp. 520-528, 2015.

[13] O.Betari, M. Erramdani, S. Roubi, K. Arrhioui, and S. Mbarki, "Model transformations in the MOF meta-modeling architecture: from UML to codeIgniter PHP framework," Europe and MENA Cooperation Advances in Information and Communication Technologies, vol. 520, pp. 227-234, 2016.

[14] Object Management Group (OMG), MOF 2.0 QVT. http://www.omg.org/spec/MOF/2.0/PDF

[15] A. Vaisman, E. Zimányi, Data Warehouse Systems: Design and Implementation, Springer Berlin Heidelberg, 2014.

[16] W. H. Inmon, Building the Data Warehouse, John Wiley & Sons, 2002.

[17] E. Malinowski, E. Zimányi, Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications, Springer-Verlag Berlin Heidelberg, 2008.

[18] S.L. Mora, J. Trujillo, I.Y. Song, "A UML Profile for Multidimensional Modeling in Data Warehouses," in Data & Knowledge Engineering, 59 (3), pp. 725-769, 2006.

[19] I.Arrassen, A.Meziane, R.Sbai, M.Erramdani, "QVT Transformation by Modelling - From UML Model to MD Model," in International Journal of Advanced Computer Science and Applications (IJACSA), 2 (5), pp. 7-14, 2011.

[20] R. . S. Kaplan, D. P. Norton, "The Balanced Scorecard – Measures that Drive Performance," in Harvard Business Review (HBR), 69 (1), pp. 71-79, 1992.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Design of a Smart Model For Geolocalisation and E-commerce in the Semantic Web

Anas Bennani, Samira Khoulji, Kerkeb Mohamed Larbi,
Information Systems Engineering Research Group, Dept. Telecommunications.
National School of Applied Science of Tetouan,
Abdelmalek Essaadi University, Morocco;
anasbenn@gmail.com; khouljisamira@gmail.com; kerkebml@gmail.com

ABSTRACT

Currently, e-commerce has become a pillar of the economy; there are huge growths of websites that offer different products to sell.This variety of web portals and products requires intelligent and autonomous operation to successfully the buying / selling process.The semantic web which is an initiation of the inventor of the web Tim Berners-lee proposes advanced technical solutions that act on the organization and the structuration of the data through ontologies.

In this article we will discuss the subject of e-commerce in company with the area of geolocation. This document proposes a new solution allowing taking advantage of the geographical location of a customer to be located in relation to the stores that surround it.

Keywords: e-commerce, Semantic Web, GEORSS Model, GEO OWL,OGC GML,Ontlogie FOAF, GoodRelations. Geographic information.

1 Introduction

In a world where consumption has become a way of life, the variety of products exposed to sales, the number of shops and supermarkets is increasing, implies the need to put in place computer tools allowing the establishment of a " An optimized policy that completes the whole process: from the presentation of the products to the payment. Actors responding in the field of e-commerce encounter difficulties. On the one hand, the seller must clearly identify his products to customers by specifying all useful information such as price, product description, image, product reviews by other customers who have purchased it.On the other hand, the customer would like to find the product wishing to buy quickly by offering him the necessary information to complete its purchasing operation. It should be noted that these products sold presented on websites, and in the majority of cases, sellers have a real place with an addressin a geographical area, they can sell either online or in their local.

The idea of this research component is to add to the e-commerce process the notion of geographical location. For example, a tourist who finds himself alone in a city that has never visited before, he would like to locate himself in his entourage and find a shop if he exists that sells a particular product, then we are in front of a repetitive scenario for each new visitor, The thing that the ambitious has given us to

DOI: 10.14738/tmlai.54.3182
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3182

contribute to improve this problem by proposing a new model allowing the tourism sector or others to benefit.

The word « semantic » in the semantic web domain's specifies the meaning of data on the Web. It is can be discovered by people and also by software. On the other hand, before the appearance of the semantic notion's, a person that read web pages and who create special software those just can infere the senses of the data. The « semantic web » significates an idea witch the software, machines and the person can find, read, understand and use the data on the World Wide Web to achieve useful goals for users. In short, the Semantic Web is supposed to make data situated anywhere on the web accessible and comprehensible by people and by the machine. [1]

It is in this perspective that we are interested in the service of the internet which is the web and more precisely the semantic web which is one of the main emerging actors in this field. Its role is to structure the data via logical links between them.

In parallel we are interested in geographical information (GI) which has become a part of our daily life, especially as we will use it in the GoodRelations ontology which is the vocabulary of E-commerce. The GI is used as an example to decide the location of a new store or store for the benefit of a company or a trader. More recently, the growth and use of (GI) was exposed by The implementation of Web-based tools such as Google Earth and the creation of GI resources by volunteers, an example being OpenStreetMap, a web-based cartographic resource. The use of geographic information has increased enormously, but many users will be largely unaware of the role it plays in their activities. This is because the GI is very rarely an objective in itself;Rather, the scenario on which the topic of interest is remarkable is normally formed.[2]

To create data through the semantic web, we mainly use an ontology[3]. It is a formalization of a conceptualization; it is definition most popular for an ontology. In terms of our semantic web domain, W3C defined an ontology as follows: « An ontology formally defines a common set of terms that are used to describe and represent a domain . . . An ontology defines the terms used to describe and represent an area of knowledge. [4]

GeoRSS is a variety of RSS that encodes the geographic location. RSS (Syndication Really Simple or RDF Site Summary) is a web data feed that can be in the form of news reports, quotes or blog topics[5]. Geo OWL presents an ontology that suits the GeoRSS function model. This model will be detailed in a chapter of this document[6].

2 Existing Ranking Methods

2.1 Ontology for E-commerce

GoodRelations is an ontology dedicated to electronic commerce. It is the strongest vocabulary used to describe products and services. The internet's user can benefit from all the advantages of GoodRelations via the use of his computer, it's able to extract and present the desired product information easily. This e-commerce vocabulary is written by adding codes on web content.[7]

To represent the e-commerce scenario, the Agent-Promise-Object-Compensation model is used. This one is done by 5 entities: An agent is a person or an organization, a promise or offer is a procedure to transfer

some rights on some object or to provide some service, an object represents product or service, a certain compensation or an amount of money, It is made by the agent and related to the object or service. A location is considered a fifth entity, it determines the place of avaibility of an offer.[7]

The five classes of GoodRelationsare: Gr: BusinessEntity coincides with the agent. Gr: Offering blends with an offer.Gr:PriceSpecification represents the compensation, Gr:ProductOrService for the object or service, Gr: Location represents the five entity. [7]

We are interested in the part of Location whose object of our research. The following image shows the class gr: location and its parameters. [7]

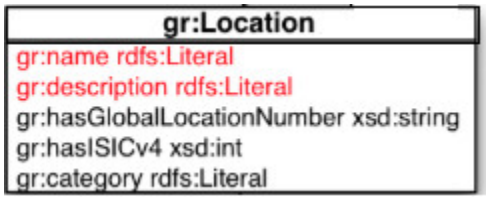
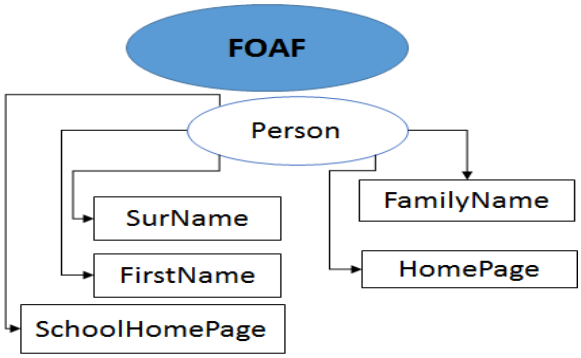


Figure 1. Location Class of GoodRelations

2.2 Ontology for Person

FOAF ontology “Friend of a Friend” describes people and the relationships they have with each other. It is one of the most popular ontologies in the semantic web. It represents personal information including name, mailbox, home page URL, friends, etc.[8]

FOAF is a dictionary of words that makes a link between people that is used in structured data [8]. It serves to publish web pages for people, groups, companies, etc. This project was founded by Dan Brichleyand Libby Miller. It enriches information about people and their relationships. In this sense, FoaF plays an important role in the creation of information systems that support online communities [9]. Among the [8]personal information that FOAF supports, we can say: FirstName, LastName, SurName, workInfoHomepage,..etc.



2.3 GeoRSS Model

The figure 3 shows the UML design of the GeoRSS model. The left part of the figure shows GeoRSS, the one on the right represents the "external" content that this model uses for its description. The association "Where" links geometry and certain external content. To be able to use this model which is an abstract concept, it must be expressed in XML, RDF, etc. This is called serialization [10].The GeoRSS schema

supports four types of geographic locations: point, line, polygon and Box, they are encoded through a Latitude and Longitude literal string called GeoRSS simple, or through a more robust way using GML called GeoRSSGML[5].

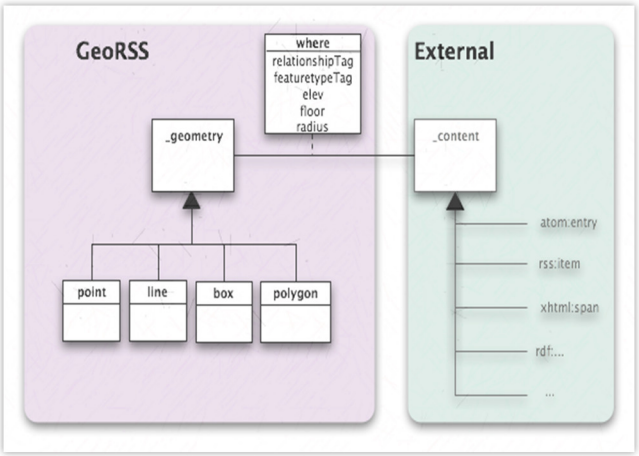


Figure 3. GeoRSS Model

In this document, we will focus on two types: Point and Box that we will define. We will use them in the next chapter.

A point is defined by a pair of coordinates, it contains two values separated by a space:Latitude and after longitude.

A box is defined in two points separated by a space, the first point is the lower corner, which must be far from the second which represents the upper corner, by definition the frame does not contain the North and South poles in its boundary. A box is used to delimit in a rounded way a zone inside.

Figure 4. Delimitation of the box type zone

2.4 GeoRSS and FOAF

This part of this document shows a basic example that combine the setting of the location of a person through the Geo and FOAF vocabularies. [11]

```
<rdf:RDFxmlns="http://xmlns.com/foaf/0.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:geo="http://www.georss.org/georss/">
<Person>
<name> BENNANI ANAS</name>
<geo:point>35.575.37</geo:point>
</Person>
</rdf:RDF>
```

3 Georss And Goodrelations

The vocabulary Schema.org is recommended for encoding the geo position, this time the coding below shows how to combine geo with GoodRelations ontology (GR) while using shema.org [12].

```
@prefix s: <http://schema.org/>
foo:posagr:Location;
gr:name "Hepp's Bagel Restaurant Munich - Bagel Street";
s:address [ a s:PostalAddress;
s:streetAddress "Bagel Street 1234";
s:postalCode "12345";
s:addressLocality "Munich, Germany" ];
s:latitude 45.75;
s:longitude 49.98;
s:telephone "+49-89-12345678-0".
```

3.1 A New GeoRSS model approach

The objective of the model that we implement in this paragraph is to give a tourist or other person connected through a geographic information system the possibility of positioning in a region. And this is affordable by adding a parameter that indicates the desired margin. This parameter is represented in decimal degrees. And therefore all GPS coordinates are expressed by the same unit.

As a legitimate scenario for this approach, a tourist doesn't know enough about the interesting areas of the city she visits, he would like to know if a perfume store or clothing is in his entourage at a precise distance. So this model will answer this problem by representing the differents steps From conception to implementation.

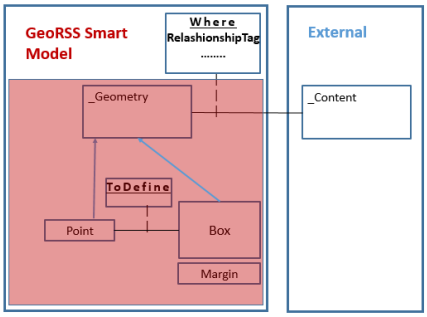


Figure 5. GeoRSSSmart Model

Let us say that the latitude position of the person is expressed by PLat and its positition longitude is PLong. The distance will be expressed by the parameter Margin, the Box propertyas shown in the figure below will have the following values: PLat-Margin PLong-Margin representing lower corner, and PLat + Margin PLong + Margin representing upper corner. As you notice the Box property will be implemented in an automatic way, this algorithm makes it easier for the person to know the main places that surround them and he can visit them and take advantage of their opportunities.

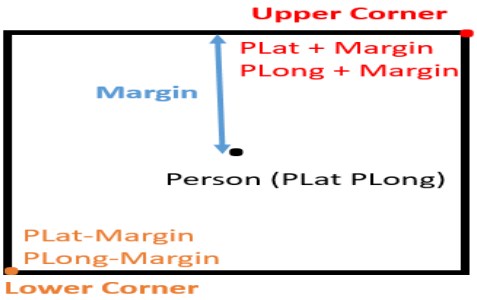


Figure 6.Definition of the Box property area

3.2 RDF Graph For Model

In this section, we will present the RDF graph of the new GEORSS model.

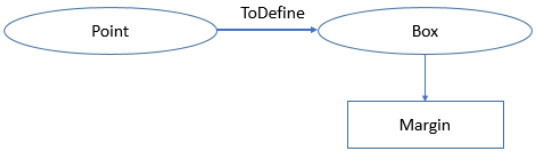


Figure 7. RDF Graph ForGeoRSS Smart Model

3.3 Implementation of a new model in a semantical way

To represent the knowledge built on the basis of the RDF model above, we opt for the OWL language: Web Ontology Language.[13]

```
<!—- Class Definition -- >
<OWL:Classrdf: ID =”_Geometry”/>
<OWL:Classrdf: ID =”Point”/>
<rdfs:SubClassOfrdf:resource=”#_Geometry”/>
</OWL>
<OWL:Classrdf: ID =”Box”/>
<rdfs:SubClassOfrdf:resource=”#_Geometry”/>
</OWL>
<!—- Object Proprieties -- >
<OWL:ObjectPropertyrdf: ID=”ToDefine”>
<rdfs:domainrdf:resource=”#Point”/>
<rdfs:rangerdf:resource=”#Box”/>
</OWL:ObjectProperty>
```

```
<!—- Data Type Proprieties -- >
<OWL:DataTypePropertyrdf:ID=”Margin”>
<rdfs:domainrdf:resource=”#Point”/>
<rdfs:rangerdf:resource=”#Box”/>
<rdfs:rangerdf:resource=”_&xsd;float”/>
</OWL:DataTypeProperty>
```

4 Conclusion

In this research component, we focus on the semantic web applied to e-commercein association with geographic information. The goal is to make the electronic commerce smarter allowing to any Internet user to carry out automatically a complex commercial operation with a minimum of interventions.

The internet users become more attached to technology, which causes their behavior change quickly, and the addition of the geo-location component; which has a direct relationship with Internet users; in the field of e-commerce participates in its progress in an accelerated manner.

REFERENCES

[1] THOMAS B.PASSIN, “Explorer’s Guide of the Semantic Web”,Manning Publication Co, ISBN 1-932394-20-6 p3 ,2004, pp. 3.

[2] GLEN HART,CATHERINE DOLBEAR, Ebook “Linked Data A Geographic Perspective”, ISBN: 13: 978-1-4398-6997-0, 2013.

[3] GRIGORIS ANTONIOU, FRANK VAN HARMELEN, “A semantic web primer”, FlteBook,The MIT Press Cambridge Massachusetts, London, England 2004.

[4] LIYANG YU, “A developer Guide to the semantic web”, Springer, ISBN : 9783642159695,pp. 136,137, 2011.

[5] LEO ORBST,TERRY JANSSEN,WERNER CEUSTERS, “Ontologies and Semantic Technologies for Intelligence”, IOS Press, ISBN: 9781607505808, p 196, 2010.

[6] Official web site: https://www.w3.org/2005/Incubator/geo/XGR-geo-20071023/

[7] Official web site Martin Hepp, GoodRelations Official Web Site: http://www.heppnetz.de/projects/goodrelations/

[8] Friend of a Firend Project Official web site,Http://www.foaf-project.org/

[9] LI DING, LINA ZHOU, TIM FININ, ANUPAM JOSHI, “How the Semantic Web is Being Used:An Analysis of FOAF Documents” System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on (2005), pp. 113c-113c

[10] Official web site http://www.georss.org/model.html

[11] https://www.w3.org/2005/Incubator/geo/XGR-geo-20071023/#examples

[12] http://wiki.goodrelations-vocabulary.org/Documentation/Geo_data

[13] STEPHEN CRANEFIELD, “UML and the Semantic Web”, University of Otago, February 2001

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Semantic Web Service Discovery Framework using Multi-Agents System and NLP Techniques

A. Essayah, M. Youssfi, E. Illoussamen, K. Mansouri, M. Qbadou
Laboratory: Signals, Distributed Systems and Arti¬cial Intelligence (SSDIA)
ENSET Mohammedia, Hassan II University, Casablanca.
Bd Hassan II, Mohammedia, Maroc.
essayah.ahmed@gmail.com, med@youssfi.net, illous@hotmail.com, khmansouri@hotmail.com, qbmedn7@gmail.com

ABSTRACT

As a consequently increase amount of web services available on the web, automatic discovery presents a great challenge, in order to satisfy this requirement, semantic discovery approaches based on ontologies have been developed. However, end-users do not have knowledge about semantic web languages to express their requirements.

In this paper, we propose an automatic discovery framework based on multi-agent systems and natural language processing techniques to match a user request. The framework allows semantic matching through a set of semantic web services in order to enhance the accuracy of the discovery pattern, and to find a relevant match to the user request composed of keywords written in natural language. The use of multi-agent systems provide us the possibility to decrease the run-time by parallelizing simultaneous tasks, also, to implement various natural language processing techniques and matchmaking algorithms, and finally, they allow us to measure the users satisfaction through their behaviors dashboard analyze.

Keywords—Semantic web Service, Ontology, web service composition, functional and non-functional properties, Distributed information systems, natural language processing.

1 Introduction

Nowadays, with the emergence of web services, semantic web and service oriented architecture (SOA)[21], the number of web services over internet is rising, a demand increases for an automatic web service discovery framework that can provides high relevant services to the user requirements expressed using queries specified by functional and non-functional properties [19].

Web services are implemented through standard technologies: WSDL [14][25], UDDI [8], and SOAP[2]. These technologies facilitate the description, discovery, and communication between services. However, this basic infrastructure encode only a syntactic representation of what is expected by, and returned from, a service (its inputs and outputs), and does not allow to automate related web service management tasks such as discovery, selection, composition, invocation and orchestration.

DOI: 10.14738/tmlai.54.3183
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3183

Web service discovery [10] is the process of searching relevant services to a user query. This process usually consists of matching a user request to a set of web service descriptions, in order to retrieve highly appropriate services to the user requirements.

Many syntactic and semantic web service discovery approaches have been proposed [16][1][13][15]. However, they present various limitations. First, some proposed frameworks require that the both of the web service provider and requester use the same ontology domain to allow the sematic matchmaker to measure the match degree between the user query and the web service description. Also, most of domain ontologies are mono-language. That means other web services described by different description languages will not be considered. Another limitation of some proposed approaches is that they require the requestor to be familiar with semantic web languages and express his requirements using a specific semantic description language such as WSMO [11], WSDL [24], OWL or OWL-S [7].

We address these limitations by providing a new framework for discovery of semantically enriched web services, the proposed framework allows end-user to search web services described using several semantic web description languages, through a request expressed in natural language keywords. The discovery process of our framework can be summarized as follows:

Understanding user requirements using natural language techniques [23] such as word sense disambiguation [22], part-of-speech tagging [4], tokenization, sentence segmentation, Stop word removal, lemmatization and Stemming to establish the user request context.

Extracting keywords from semantic or syntactic web service description to specify the web service context.

Measuring similarity degree between user request context and web service context.

Ranking the selected web services using several matchmaking algorithms considering functional and non-functional properties (e.g. price, location, service availability, Quality of Service ...)

Measuring the user satisfaction to adapt the discovery mechanism to the user preferences for later requests.

Multi-agent systems [5] provide a new method to analyze, describe and implement sophisticated applications that need to distribute tasks between autonomous entities. The use of Multi-agent system allows us to adopt various natural language processing techniques such as PoS [4], WSD [22] … and discover web services described in different web service description languages like WSDL [11], WSMO [11], OWL-S [7] …, with a minimum adaptation for Worker Agents. Agents ensure a good portability because they can be deployed in several types of devices including mobile devices.

The remaining of this paper is organized as follows. In section 2 we provide an overview of the related work performed in the area of semantic web service discovery, whereas in section 3, we provide some background knowledge about technologies that are integrated in our framework, namely NLP techniques, web service description languages and Multi-agent systems. In section 4 we present an overview of the proposed framework then we discuss in details its architecture. Finally, section 5 concludes the paper.

2 Related Works

Many research have been made to create a web service discovery framework. Generally, all the proposed approaches are based on syntactic or semantic matchmaking algorithms.

M. Paolucci, T. Kawamura, T.R. Payne, and K. Sycara [16] proposed a semantic matchmaking algorithm consisting of a similarity degree obtained by matching the user request inputs and outputs to the web service inputs and outputs. Four different matching degrees are possible: Exact Match (When the user request inputs/outputs match exactly the web service inputs/outputs), Plugging Match (When the web service inputs/outputs plug the user request inputs/outputs), Subsumed Match (When the user request inputs/output subsume the web service inputs/outputs), and otherwise, a Fail degree is generated.

A. Bener, V. Ozadali, and E. Ilhan [1] extended Paolucci's matchmaking algorithm by proposing an algorithm that performs as well as inputs/outputs matching, preconditions, effects and service context. J.M. García, D. Ruiz, A. Ruiz-Cortés [13] proposed to improve the run-time by adding a preprocessing stage to discard all web services that not contain the concepts expressed by the user request. Using SPARQL this preprocessing mechanism improve the run-time, but it has a negative influence in term of precision.

A knowledge based approach for semantic web services discovery have been presented in [15], the mechanism is based on knowledge and multi-level search for semantic web services discovery. Which an intelligent broker utilizes a knowledge based system and multi-level search to overcome the drawbacks of syntactic and semantic discovery. The multi levels of service discovery allows them to improve the probability of accurate matching.

Our proposed framework considers all the end-users as a non-technicians, we extract concepts from a keyword-base user request using NLP techniques and we define automatically the user request and the web service context, all cited algorithms are effective for single web service description language, instead of from our approach that can be used to discover services described in various web service description languages such as OWL-S, WSMO, WSDL,…

3 Background

In this section we briefly describe some concepts that are integrated in our proposed framework, we start by presenting some natural language processing techniques used to process a requestor requirements expressed using a human language, Then, we define some semantic web concepts and languages used to describe syntactically and semantically a web service, and finally we finish this section by presenting an overview about multi-agent systems.

3.1 Natural language processing techniques (NLP).

Refers to the use and ability of computers to process and understand sentences expressed in a human (natural) language such as English, NLP [23] is a growing area of research with several applications in different fields. Technically our proposed framework employs a set of natural language processing techniques which are listed below:

Tokenizing[4]:Word tokenizing is an important part of text processing, every natural language needs to normalize the text and chop it up into pieces called tokens, text tokenization may also ignores some

characters such as punctuation. For example, a sentence like “Hotel in Casablanca Morocco, with a good price” can be divided into 8 tokens {“Hotel”, “in”,“Casablanca”, “Morocco”, “with”, “a”, “good”,“price”}.

Stop word removal: stop word is a predefined set of words from a language that are excluded from the vocabulary because they don’t give a relevant semantic information. An example of stop words list is shown in Figure 1.The last example will be composed of five tokens after applying Reuters-RCV1 stop word list {“Hotel”, “Casablanca”, “Morocco”, “good”, “price”}.

a	an	and	are	as	at	be	by	for	from
has	he	in	is	it	its	of	on	that	the
to	was	were	will	with					

Figure.1. Stop list of 25 semantically non-selective words which are common in Reuters-RCV1

Part-of-speech Tagging[23]: PoS tagging is the process of tagging words based on their context with its corresponding part-of-speech (PoS). Many PoS Tagger are available such as Brill’s rule-Based PoS Tagger [4] from natural language toolkit library (NLTK) [3].

Stemming and lemmatization: the aim of stemming and lemmatization process is to reduce tokens into their stem form called lemma. Depending stemmers/lemmatizers tokens stemming and lemmatization can be done various techniques such as, Terms normalizing (e.g. are, is, am = be), equivalence classes definition, asymmetric expansion (e.g. Hotel, hostel, motel, resort = hotel), affixes chopping (e.g. development, developments, developing, developer = develop). Many words lemmatizers/stemmers are available, the most common for English language is Porter Stemmer [20] and WordNet lemmatizer from NLTK library [3].

Word sense disambiguation [22]: after tokens lemmatization a word sense disambiguation process takes place to identify which meaning of a polysemy lemma is employed in a sentence. WSD algorithms use the sentence’s context and previous disambiguated words to calculate a similarity score.in our case, word sense disambiguation will be applied to the set of lemmas extracted from the user request and from the web service description, producing two sets of two sets of disambiguates meanings that can be used in the discovery and matchmaking algorithms.

3.2 Web service description languages

To promote the automation of web service management tasks, a number of different semantic languages have been created such as, WSMO [11] and OWL-S [7], to provide a more powerful method for web service discovery, selection, composition, invocation and orchestration by describing functionalities of services in a machine interpretable form.

WSDL (Web Service Description Language)[14]: is a W3C (World Wide Web Consortium) standard for syntactic web service description, WSDL 2.0[24] used to describe the web service interface and the message exchange patterns (Structure of SOAP messages).

OWL-S (Ontology Web Language for Services)[7]: formerly DAML-S, is an ontologies-based web service description language, in OWL-S each service is described using three XML-Based parts which are: Service Profiles (to describe the web service capabilities), Service Model (describing web service behaviors) and Service Grounding (to provide details about the web service invocation).

WSMO (Web Service Modelling Ontology)[11]: provides a semantic modelling language for web services capabilities description. In WSMO, a web service is described using four core elements, which are: Ontologies, Goals, Web services and Mediators.

UDDI (Universal Description Discovery and Integration)[14]: is a standard web service registry, that allows services providers through an API to publish his services, and services requesters to search for desired web services. UDDI offers a syntactic description of its content. However, it does not present semantic models to understand the user requests and web service capabilities.

3.3 MULTI-AGENT systems

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Multi-agent system (MAS) [5] is a collection of interacting agents that are used to solve complex and distributed problems. Agents are loosely coupled which allows adding new functions or removing existing features easily by creating or destroying agent without affection the whole system structure. In our case, the following features of MAS are especially interesting:

Goal orientedness: An agent is able to handle complex tasks, it should decide in which order and in which way the task should be processed. That feature allows us to adopt various natural language processing techniques and discover web services described in different web service description languages with a minimum adaptation for Worker Agents.

Autonomy: Agents are highly autonomous, they operate without human or other agents' intervention. Which allows us to add and remove various functionalities to our framework without affection the system.

Portability: Agents can be deployed in various devices that are supporting JADE/LEAP [5], such as mobile devices, personal computers, servers.

4 The Proposed Framework

4.1 Semantic Web Service Discovery And Composition Framework

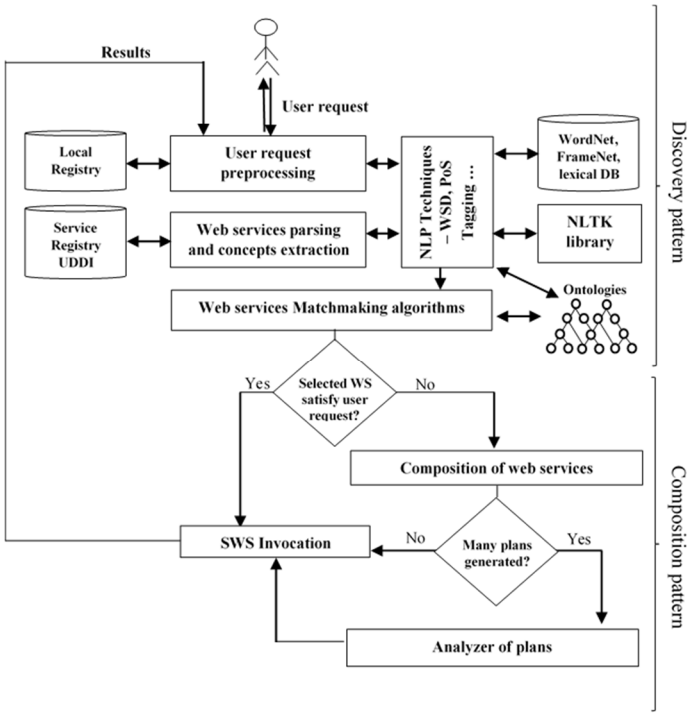


Figure.2. Web service discovery & composition framework global architecture.

Our proposed framework shown in figure 2 aims to discover web services using various web service discovery techniques in order respond to a keyword-based user request expressed using natural languages. When no selected web service matches the user requirements, a service composition process take place to create new functionality by combining the features offered by other existing services. Our proposed framework consists of two different patterns:

- Discovery pattern: The web service discovery pattern consists of keywords-based discovery process for searching web services described using semantic or syntactic languages, the discovery pattern use NLP techniques to establish a match between a user request expressed using keywords and a semantic or a syntactic web service description.
- Composition pattern: The composition mechanism take place when no selected web service meet the user requirements, Our approach based on workflow, consider a composite service as a business process. A composite service (composite plan) includes a set of atomic services as well as controls and data exchange between them. Similarly, a business process is composed of a set of elementary structured activities, as well as the execution order between them.

In this article, the focus is on the discovery pattern, the composition process will be the subject of future works.

4.2 Web Service Discovery Pattern

To satisfy the user requirements, the web service discovery pattern shown in figure 3, uses many components. First, a proxy agent which is the interface between the user and the platform, it's capable of receiving the user request, looking for a response in the local registry, communicating with intermediate agents, executing services, delivering results to the user, measuring user satisfaction and saving the relevant solution to the user request as an history in the local registry. The proxy agent can be hosted by personal computers, servers and mobile devices that are supporting JADE/LEAP [5] such as smart phones, cars, smartwatch and all other wearable devices. Second, a NLP container that allows keywords-based discovery process for searching web services described using semantic or syntactic languages. And finally, a discovery container that uses various web service description languages parsers to extract web service descriptions from the web service register, and different matchmaking algorithms to match between a user request expressed using keywords and a semantic or a syntactic web service description.

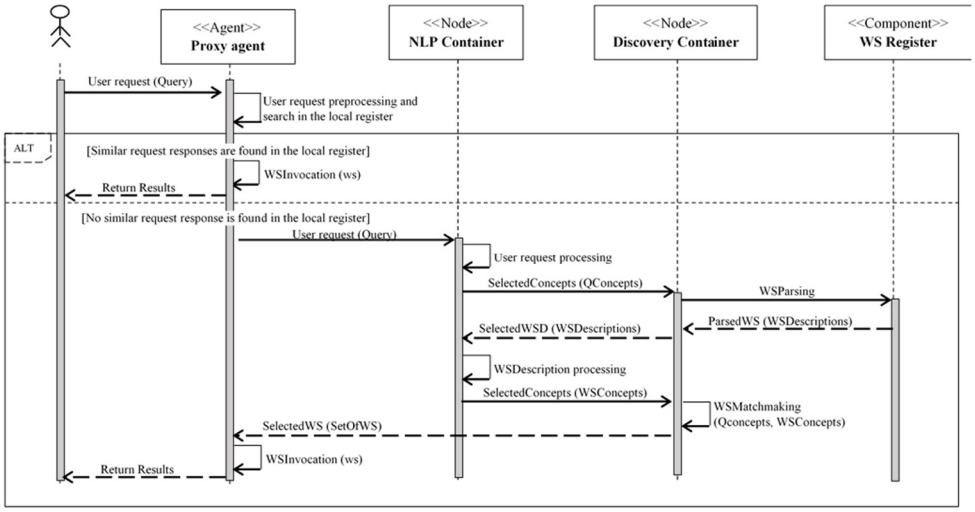


Figure.3. Web service discovery pattern sequence diagram

4.2.1 Natural language processing (NLP) Container:

A service requestor express its requirements using a collection of words. Because a word can have various senses (e.g., pen can be used to mean either an instrument for writing or drawing with ink or a female swan).

NLP container as shown in figure 4, is used to process the user request and the web service descriptions and extract the context and the meanings. The NLP mechanism start by receiving a user query from the proxy agent, the intermediate agent, is the interface between the NLP Container agents and the other system's components. The intermediate agent initiates the user request and the web service description processing by sending a tokenizing request to the NLTK library, to normalize the text and chop it up into pieces called tokens. After receiving tokens the intermediate agent calls a PoS tagging method to tag tokens based on their context with its corresponding part-of-speech (PoS), then we remove the stop words. The intermediate agent then creates WAn worker agents corresponding the number of tokens, each worker agent send a stemming request to the NLTK library to reduce token into its stem form called lemma. In order to identify which meaning of a polysemy lemma is employed, worker agents invoke a

word sense disambiguation method, selected concepts are send to the intermediate agent to be delivered as a set of user query concepts (QConcepts) or web service description concepts (WSConcepts) to the discovery container to start the discovery process.

4.2.2 • Discovery Container:

The web service discovery container shown in figure 5, aims to select web services from the service registry, parse and read web service descriptions to extract functional and non-functional concepts and attributes and finally rank selected web services using matchmaking algorithms.

After receiving a discovery request from the NLP Container, the intermediate agent start by creating different Parser agents, each parser agent use specific web service description reader and parser to be able to read and extract relevant information through web service description written in different languages. For OWL-S a parser like Sesame [18] or Jena [12] can helps in this process and for WSMO which is written using WSML parser like WSMO4J [8] can be used. Once the parsing process finish, a sets of relevant web services description related words are selected and sent to the NLP Container (figure 3) to be processed. The NLP Container define the meanings of each word by applying a various NLP techniques like word tokenizing, Stemming and lemmatization and word sense disambiguation. When the NLP mechanism is done, the discovery container receive again a set of selected user request and web services description meanings to match. The goal of matchmaking is to process the user request and the web service description concepts and measure the similarity degree between them. A ranked set of the selected web services is generated. Finally, the proxy agent invokes the selected web service and measures the user satisfaction to adapt the discovery mechanism to the user preferences for later requests.

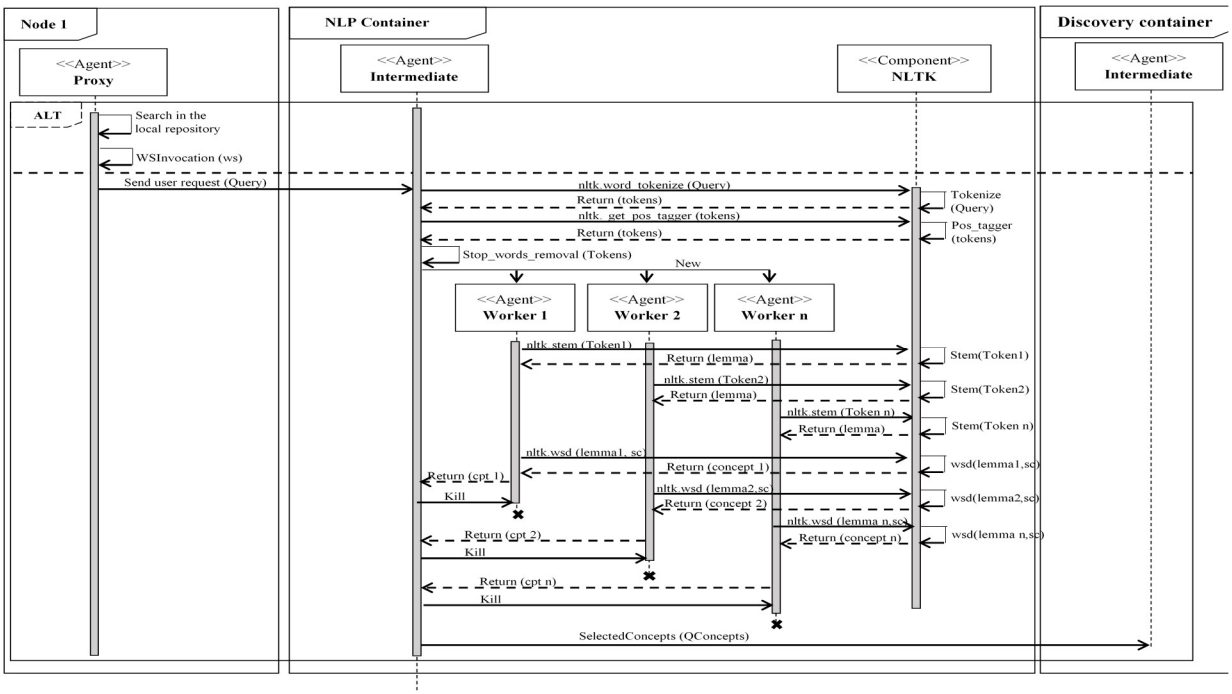


Figure.4. NLP Container sequence diagram

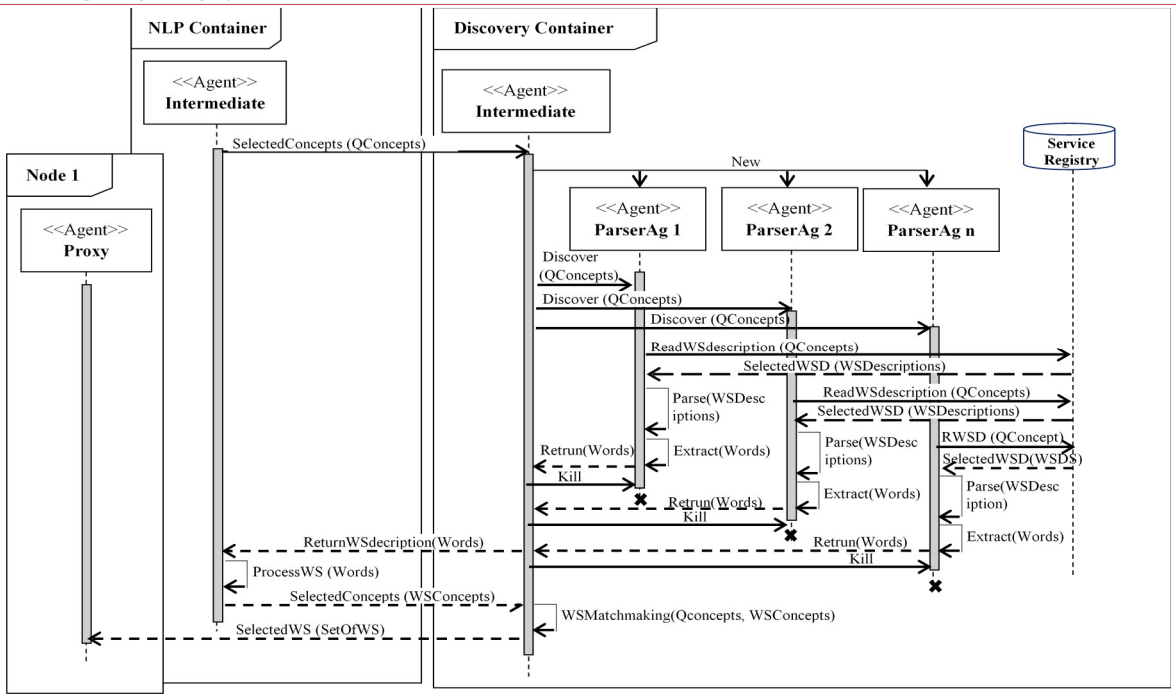


Figure 5. Discovery Container sequence diagram

5 Conclusion and Perspectives

In this paper we presented a framework that allows web service discovery based on user request expressed using human language.

Our proposed framework considers the end-user as non-technician, in fact, we don't require any knowledge related to the web service technologies, the discovery process start by applying NLP techniques to extract concepts from a keyword-base user and then defining automatically the user request and the web service context. Furthermore, our proposed system uses various web service matchmaking algorithms to find the best match to the user requirements. The use of Multi-agent system permits us to implement several web service description languages parsers and execute parallel tasks in order to improve the run-time.

To extend our proposed framework, composition mechanism is required. Future work includes more details about our composition pattern based on workflow and considering a composite service as a business process.

REFERENCES

[1] A. Bener, V. Ozadali, and E. Ilhan, Semantic matchmaker with precondition and effect matching using SWRL. Expert Systems with Applications, 36(5), 9371–9377, (2009).

[2] A. Navarro, A. da Silva, A Metamodel-based definition of a conversion mechanism between SOAP and RESTful web services, Computer Standards & Interfaces, Volume 48, Pages 49-70, ISSN 0920-5489? November 2016.

[3] Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O'Reilly Media Inc. (2009)

[4] Brill, E., 1992. A simple rule-based part of speech tagger. In: Proceedings of the Workshop on Speech and Natural Language. Association for Computational Linguistics, pp. 112–116.

[5] C. Castelfranchi, Y. Lespérance, Developing Multi-agent Systems with JADE Intelligent Agents, In Intelligent Agents VII Agent Theories Architectures and Languages, Vol. 1986,(2001)

[6] D. D. Lewis,Y. Yang, T. Rose, and Li, F. RCV1: A New Benchmark Collection for Text Categorization Research. Journal of Machine Learning Research, 5:361-397, 2004. http://www.jmlr.org/papers/volume5/lewis04a/lewis04a.pdf.

[7] D. Martin, M. Paolucci, S. McIlraith,M. Burstein,D. McDermott, D. McGuinness and al. Bringing semantics to web services: The OWL-S approach. Semantic Web Services and Web Process Composition; p 26-42, 2005.

[8] EU IST, FIT-IT. (2008). WSMO4J API. Available from<http://wsmo4j.sourceforge.net/>.

[9] H. C. Boas, "Bilingual FrameNet Dictionaries for Machine Translation", in Proceedings of the Third International Conference on Language Resources and Evaluation, Las Palmas, vol. IV, pp. 1364-1371. 2002.

[10] H. Cancela, A. Cuadros-Vargas, A. D. Renzis, M. Garriga, A. Flores, A. Cechich, A. Zunino, CLEI 2015, the XLI Latin American Computing ConferenceCase-based Reasoning for Web Service Discovery and Selection, Electronic Notes in Theoretical Computer Science, Volume 321, Pages 89-112, 2016.

[11] Hai H. Wang, Nick Gibbins, Terry R. Payne, Domenico Redavid, A formal model of the Semantic Web Service Ontology (WSMO), Information Systems, Volume 37, Issue 1, Pages 33-60, March 2012.

[12] HP Labs Semantic Web, (2009). Jena. Available from<http://jena.sourceforge.net/>.

[13] J.M. García, D. Ruiz, A. Ruiz-CortésImproving semantic web services discovery using Sparql-based repository filteringWeb Semantics: Science, Services and Agents on the World Wide Web, 17, pp. 12–24 (2012).

[14] M. B. Juric, A. Sasa, B. Brumen, I. Rozman, WSDL and UDDI extensions for version support in web services, Journal of Systems and Software, Volume 82, Issue 8, Pages 1326-1343, August 2009.

[15] M. El Kholy and A. Elfatatry, "Intelligent broker a knowledge based approach for semantic web services discovery," Evaluation of Novel Approaches to Software Engineering (ENASE), 2015 International Conference on, Barcelona, pp. 39-44, 2015.

[16] M. Paolucci, T. Kawamura, T.R. Payne, K. Sycara, Semantic matching of web services capabilitiesThe Semantic Web–ISWC 2002, pp. 333–347, Springer (2002)

[17] OASIS, 2004. UDDI Version 3.0.2, UDDI Spec Technical Committee.

[18] openRDF.org. (2009). Sesame. Available from<http://www.openrdf.org/>.

[19] P. Li, M. Comerio, A. Maurino and F. D. Paoli, "Advanced Non-functional Property Evaluation of Web Services," Web Services, 2009. ECOWS '09. Seventh IEEE European Conference on, Eindhoven, 2009.

[20] Porter, Martin F. An algorithm for suffix stripping. Program 14 (3): 130-137.1980.

[21] R. J. Rabelo, O. Noran and P. Bernus, "Towards the Next Generation Service Oriented Enterprise Architecture," 2015 IEEE 19th International Enterprise Distributed Object Computing Workshop, Adelaide, SA, 2015, pp. 91-100.

[22] R. Navigli, & P. Velardi. Structural semantic interconnections: A knowledgebased approach to word sense disambiguation.IEEE transactions on pattern analysis and machine intelligence (Vol. 27, pp. 1075–1086). IEEE Computer Society (2005).

[23] Venkat N. Gudivada, Dhana Rao, Vijay V. Raghavan, Chapter 9 - Big Data Driven Natural Language Processing Research and Applications, In: Venu Govindaraju, Vijay V. Raghavan and C.R. Rao, Editor(s), Handbook of Statistics, Elsevier, Volume 33, Pages 203-238, 2015.

[24] W3C, 2007. Web Services Description Language (WSDL) Version 2.0, W3C Recommendation.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Digitizing Human Sciences to Determine the Individual's Personality Based on Facial Emotions Recognition

S.Bourekkadi, S.Khoulji, O.Omari, K.Slimani, S.Chouya, M.L.Kerkeb
Information System Engineering Group, ENSA, Abdelmalek Essaadi University, Tetuan, Morocco.
slimani.uit@gmail.com

ABSTRACT

Informatics is a science that develops and opens up to a group of other sciences. The current researches have been established for multiple partnerships between informatics and the rest of the sciences. This is noticed in the integration of Informatics for instance within medicine, economics, and humanities. Researchers have made computer's software that help draw the emotional state of the individual from the analysis of the movements that appear on the face. Today, we will follow the same plan, but we will not only study facial features. We will determine the personality based on other data and also face dimensions. We will rely on determining the type of personality based on the analysis of specific parts of the face such as the upper part of the front face, the middle section of the nose and the mouth, and the lower jaw section. In the current study, we used a modern scientific methodology that is manifested in the digitization of human sciences and in the analysis of the image by adopting a software product based on determining the dimensions of the face, specifically mathematical and descriptive, and then extracting the conclusions contained in the black box of the software's information. Our conclusions and results are confirmed by a group of people whom we have adopted in our experiences. Moreover, we did not only define personality but we have also made it possible for individuals to deal with four types. We dealt with many people, but we can recognize the features of their characters only after the experience and keeping in company with them for a long time. There may also be other ways through which we can take a preliminary idea of the person's character when we deal with the features of the face.

Keywords-component; IT psychology, emotion, face, personnality,

1 Introduction

We often find two people, despite their differences in nationality, language, and culture, who have the ability to take emotional attitudes toward one another, and through facial expressions, they exchange the most different feelings and sensations. Even the people of the same country, when they get in the same bus, we find them exchanging impressions through facial movements. Everyone knows that what we mentioned is true, known, and not new.

But, what about reading the face of human beings we meet for the first time. Can we judge their actions before they do them and can we judge their qualities before we see their actions? In fact, we often harshly judge people through the

shape of their face, and that is what our ancestors once called physiognomy.

DOI: 10.14738/tmlai.54.3185
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3185

This is what is done nowadays by analyzing the personality through reading the features of the face. It is of course a distinguished science that has its theorists and professors, and has theories that support its origins and methods. In fact, there is an enormous and important amount of various studies and researches that support the argument that there is a precise and direct link between the features of the face of any person and the personality of this person.

The human's personality may be mysterious sometimes, but it can be identified by some external characteristics such as hair color, body length, skin color, eye shape, or even eye color. All of these qualities can be determined by the adoption of our information system that examines the picture in logical and accurate ways.

2 Personality Analysis From the Face

We have found that each face reveals what is hidden in the person's personality and describes its owner with extreme accuracy. This for example gives some conclusions of the qualities and features enjoyed by the owner of this face such as deception, goodness, fun, generosity, love or criminality. Meanwhile, all human specifications, being positive or negative, could be read and known through a digital analysis of the features and shape of each face.

Our information system initially records the provided image of the person whom we want to analyze his/her personality. The information system frames the face, determines the position of the face on the image, and then calculates the dimensions of the face. It takes into consideration also the comparison of the length with the width. After that, it calculates the ratio of the enlargement of the given data according to mathematical relations as displayed in Figure 1 below.

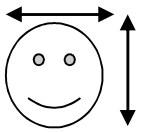
Width	The length calculated by our function	Real length	**W**idth 
x_i	$f(x_i) = 154 \times 10^{-2} x_i$	V_i	
Cases			
$f(x_i) \approx V_i$	$V_i \approx x_i + \left\lvert \frac{x - V_i}{2} \right\rvert$	$x_i \approx V_i$	Other cases by percentage $\%i = \frac{x_i \times 100}{f(x_i)}$

Figure 1. Calculated equations and conclusion

This ratio determines the type of face shown in the image, classifies it as demonstrated in Figure 2, and then connects it with a text conclusion that is appropriate to it. The following are some of the results:

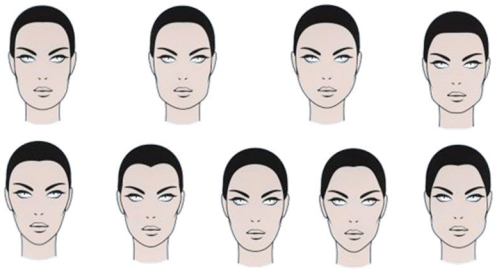


Figure 2: Face shapes

1. People who have a broad face are dominated by misbehavior and laziness accompanied with being unable to remember events, and perhaps they are not free of foolishness and stupidity.
2. People with long faces have a great deal of movement and mobility. They do all the tasks and things quickly. They have also the lightness of the spirit; however, when disturbed by others, they may turn into a very bad mood.
3. The owners of the prominent faces (which contain a prominent face bone strikingly visible) are malicious and deceitful. They also have negative qualities such as spending most of their time making conspiracies and intrigues to other people.
4. The owners of small faces are characterized by the kindness of their hearts and the purity of their intention. At the same time, they are characterized by weakness, fear, retreat and surrender due to their lack of tricks and helplessness.
5. People who have round faces are not as interested in their reasons as they are for the results, so they are positive.
6. The owners of broad foreheads are characterized by a large mind that fits with the size of their head.
7. The owners of narrow foreheads are characterized by narrow horizons, and most often look at all things in a strange superficiality.
8. People with zigzagging foreheads are often characterized by worries, problems, and many failed experiences.

3 Personality analysis by three basic spots of the face

Physiognomy Science divides this into three sections: upper, middle, and lower section. Meanwhile, we can determine the personality of the person by the dominance of one of these sections on the other two sections. In addition, this dominance determines the psychological tendencies and preparations of the person.

In this stage, the information system returns to the same image that it is working on, and after defining the system for the frame of the image, the face is divided into three important sections, (see Figure 3), namely the forehead, nose, mouth, and jaw. After that, the information system attempts to calculate the total area of the face according to a mathematical relation (see Figure 4), and then it calculates the area of each of the three sides. Later on, the system ultimately determines the two areas that were given the biggest proportion. Based on those percentages, the information system comes out with the following conclusions:

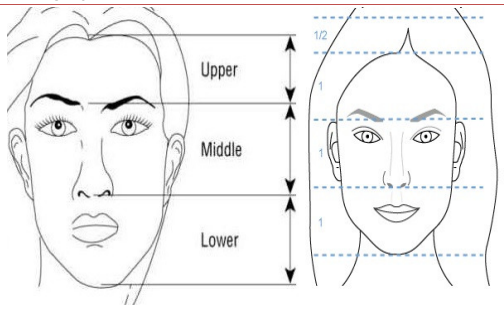


Figure 3. The important sections of the face

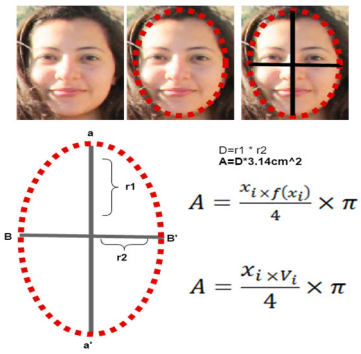


Figure 4. The mathematical relations of calculating the face area

3.1 The upper section (the forehead)

The forehead of the human's face is responsible for the extent of knowledge of human relations, social and intelligence. This is because the owner of the broad forehead has advantage of intelligence, sophisticated, and developed mental capabilities. Moreover, if the forehead is wide and rounded, then its owner has a good taste. However, if it is wide from the top, the imagination of intelligence is more dominant. In the lower section there is a manual for controlling the applied power of intelligence. Meanwhile, concerning the presence of wrinkles in the front area, researchers say that this presence of wrinkles reflects the tendency of this person to abstraction, and the search for reasons. The person has also a high ability to analyze, but in the case of the front wide and clear of wrinkles, it reflects the tendency of this person to the dreams of vigilance and inability to adapt to practical life.Meanwhile, the wide and deviated front square reflects the ability of the person toward creation and creativity, while the broad forehead in the form of the arch reflects the negativity of its owner and the infertility of his ideas.

3.2 The middle section (nose and mouth)

A person with a large nose and mouth reflects the emotional capacity and momentum joy by that person.This reflectsalsothat the person with an emotional and advanced mood, which makes him/her inclined to emotions, has a warm and calm life in addition to its superior ability to confront, and these qualities are confirmed by the presence of long and huge nose.On the other hand, the sharp nose signifies its nervousness and violence, and vice versa. This means also that the small nose and mouth and their shape indicate emotional coldness and social closure, while the owner of the delicate nose indicates his/her deep coolness and the ability to control high emotions.

3.3 Lower section (jaw)

The amplitude of the jaw and mouth reflects great physical abilities and endurance, and can control the person's emotions. The accuracy of the jaw reflects inactivity, negativity, weakness, and hesitation. It also shows the lack of diplomacy in dealing with people. On the other hand, the owner of the small mouth and lower sides reflect unhappiness and depression.

Finally, after reviewing the three sections and their descriptions, it is possible to say that few people are characterized by a clear dominance of one of these sections. We often find ourselves in front of many faces that share control between two sections at the expense of the other. It often happens that this matter is distributed over the three sections but in varying proportions.

4 Personality analysis by Eye Color

Many people want to recognize their personalities, whether through their eye color, hair color, dress, eye shape or food...etc. In this part, the information system will identify the user's eye color, by framing the eye side, and accompanying this step by presenting a short text that gives the result of the analysis. In this section, we will see the personality according to eye color.

The information system is able to determine the color of the eyes. At this stage, the information system frames the eye side and frames the center of the eye, and finally the eye color (See figure 5).

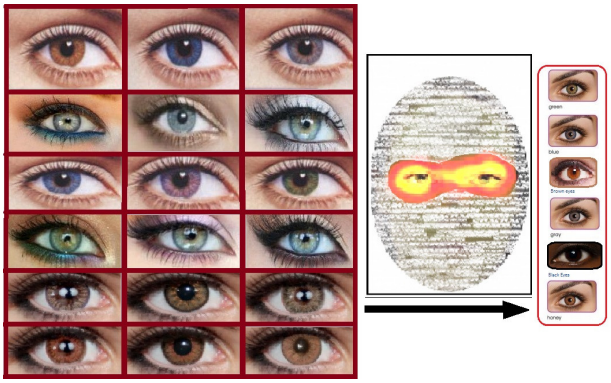


Figure 5. The different color of the eyes

This identifies some colors defined by the system and then connects the eye color to what it symbolizes according to psychology to become a digital psychologist. These are some of the results:

Green eyes

Green eyes often like to help others, and also are characterized by harshness and hardness. Green eyes also have a bold and strong personality, but sometimes they are selfish, and sometimes passionate and affectionate too.

Blue eyes

They treat people with tenderness and kindness. They also impose their attitudes without touching the other's feelings. Moreover,people with blue eyes have a great connection to what they have and they are also optimistic.

Brown eyes

People with brown eyes often have tender and emotional hearts. Moreover, the browner the eyes of the person are, the more that person showshis/her tenderness. In addition, people with brown eyesare also ambitious, keen to be calm, and they deal with people politely.

Grey eyes

The grey eyes' personalities are divided into two parts, either calm and generous, or harsh and violent. People with grey eyes often search to be calm but rarely find it. They are also very observant and there is no place for emotions in their lives.

Black Eyes

Those people are social, but frequently jealous. When they become angry, they lose control of themselves. They love to help and meet people's needs even at their expense. They also have pure tender feelings.

Honey Eyes

A personality that is not honest with itself although it is good-hearted, self-reliant, calm, and does not rush to make decisions. It is controlled by the mind and not guided by passion. It is also discrete despite its tendency to curiosity and intelligent personality.

In fact, we can compare the eye with a mirror that reflects what is within us and reflects our personalities and ourselves without knowing. This is closely linked to our personality, how we behave, what are our methods, and what are our ethics. As some scientists and researchers of psychology have concluded, there is continuity between our personalities and the eyes.

5 Conclusion

The discussion on this subject is a long and interesting inspiration with its delicate and complicated branches. Progression in the collection of knowledge on that matter requires continuous studies and research, but this is what we can show in for the limited and available space. We hope that the coming researches will examine ways in which we can help humanize some information systems that will help to sensitize some negative feelings and make them positive.

All in all, a human's personality is only understood by its owner even though there are many analyses of personalities, but this does not mean not to see them.

The discussion on this subject is a long and interesting inspiration with its delicate and complicated branches. Progression in the collection of knowledge on that matter requires continuous studies and research, but this is what we can show in for the limited and available space. We hope that the coming researches will examine ways in which we can help humanize some information systems that will help to sensitize some negative feelings and make them positive.

All in all, a human's personality is only understood by its owner even though there are many analyses of personalities, but this does not mean not to see them.

ACKNOWLEDGMENT

We would like to express our gratitude and appreciation to Pr. Azzouz Ettoussi and Souhail Bourekkadi for their great encouragement, moral supports and their guidance, and also for their help and their comments that greatly improved the manuscript.

This work was supported by the grant of National Center for Scientific and Technical Research (CNRST-Morocco): No. 6UAE2015 and 757UITfor the first and the fourth authors.

REFERENCES

[1] Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. Journal of educational psychology, 85(4), 571.

[2] Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. Psychology in the Schools, 30(1), 79-90.

[3] Bideaud, J., Houdé, O., & Pedinielli, J. L. (1993). L'homme en développement. Presses universitaires de France.

[4] KAMATE, M. Z. (2009). ETUDE DES TROUBLES PSYCHIATRIQUES CHEZ LES EMIGRES DANS LE SERVICE DE PSYCHIATRIE DU CHU POINT «G».

[5] dK.SLIMANI,S.BOUREKKADI, R.MESSOUSSI, S.KHOULJI, N.BEZAD, M.L.KERKEB “Design of an Awareness System on Sexual and Reproductive Health Based on the Recognition of Emotional Facial Expressions”, International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), March 16 Volume 4 Issue 3 , ISSN: 2321-8169, PP: 193 – 196.

[6] bRobert Escarpit, l’information et la Communication théorie générale.

[7] eHowell, D., Bestgen, Y., Yzerbyt, V., & Rogier, M. (2008). Méthodes statistiques en sciences humaines.

[8] Allais, M. (1953). La psychologie de l'homme rationnel devant le risque: la théorie et l'expérience. Journal de la Société de Statistique de Paris, 94, 47-73.

[9] Donzé, L. (2001). L’imputation des données manquantes, la technique de l’imputation multiple, les conséquences sur l’analyse des données: l’enquête 1999 KOF/ETHZ sur l’innovation. Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 137(3), 301-317.

[10] The Collected Works of John W. Tukey (ISBN 0412063212) ; (ISBN 0534051014)

[11] J W Tukey, Graphical Analysis of Multiresponse Data, K E Basford (ISBN 0849303842)

[12] Kahana RJ. Psychotherapy : models of the essential skill. In : Bibring GL (Ed.) The Teaching of Dynamic Psychiatry : A Reappraisal of the Goals and Techniques in the Teaching of Psychoanalytic Psychiatry. Madison, Conn: International Universities Press; 1968. p. 87–103.

[13] Ogden, C. K., Richards, I. A., Ranulf, S., & Cassirer, E. (1923). The Meaning of Meaning. A Study of the Influence of Language upon Thought and of the Science of Symbolism.

[14] Borillo, M. (1984). Informatique pour les sciences de l'homme: limites de la formalisation du raisonnement. Editions Mardaga.

[15] Schwartz JH, Kandel ER. Modulation of synaptic transmission : second-messenger systems. In : Kandel ER, Schwartz JH, Jessell TM (Eds.) Essentials of Neural Science and Behavior. East Norwalk, Conn: Appleton & Lange; 1995. p. 243–67.

[16] Wanlin, P. (2007). L'analyse de contenu comme méthode d'analyse qualitative d'entretiens: une comparaison entre les traitements manuels et l'utilisation de logiciels. Recherches qualitatives, 3, 243-272.

[17] Przygodzki-Lionet, N. (2014). Le témoignage en justice: les apports de la psychologie sociale et cognitive. Histoire de la justice, (1), 115-126.

[18] Heutte, J. (2011). La part du collectif dans la motivation et son impact sur le bien-être comme médiateur de la réussite des étudiants: Complémentarités et contributions entre l'autodétermination, l'auto-efficacité et l'autotélisme (Doctoral dissertation, Université de Nanterre-Paris X).

[19] Stafford, J., & Bodson, P. (2006). L'analyse multivariée avec SPSS. Puq.

[20] Tiberghien, G., & Beauvois, J. L. (2008). Domination et impérialisme en psychologie. Psychologie française, 53(2), 135-155.

[21] Christophe Clivaz, Stéphane Nahrath et Mathis Stock, "Le développement des stations touristiques dans le champ touristique mondial", Mondes du tourisme, numéro hors série "Tourisme et mondialisation", 2011.

[22] Bronckart, J. P., Bain, D. E., Schneuwly, B., Davaud, C., & Pasquier, A. (1985). Le fonctionnement des discours: un modèle psychologique et une méthode d'analyse.

[23] Jean-Marie Bouroche et Gilbert Saporta, L'Analyse des données, P. U. F, 8e édition, 2002 (ISBN 2 13 052495 8)

[24] Berriane, M. (1993, March). Le tourisme des nationaux au Maroc (une nouvelle approche du tourisme dans les pays en développement). In Annales de Géographie (pp. 131-161). Armand Colin.

[25] DRISSI HABTI, N. A. W. F. E. L. (1998). Le tourisme marocain: analyse spatio economique d un systeme soumis a la mondialisation (Doctoral dissertation, Nice).

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Wide Band Patch Antenna using Defect Ground Structure Optimizing by Genetic Algorithm

[1]Omar El Maleky, [1]Farid Ben Abdelouahab, [2]Mohammed Essaadi
[1]Laboratory Materials and Radiation, Abdelmalek Essaadi University, Tetouan, Morocco
[2]ENSIAS, Mohammed V Souissi University, Rabat, Morocco
o.elmaleky@gmail.com

ABSTRACT

The patch antenna is widely used in our life for a lot of its strong points as a low cost, easy and simple fabricated structure, and a code characteristic in radiation profile. But the narrow bandwidth is, however, the main drawback of a microstrip patch antenna. In this paper the bandwidth of a patch antenna is improved by using a technique known as defected ground structure (DGS), and the size of the antenna is optimized using genetic algorithm. As a result, we were able to optimize the size of the initial antenna proposed (50mm * 30mm * 1.55mm) and expand the bandwidth of the patch antenna. The final simulated prototype has a size of 26mm * 20mm * 1.55mm which corresponds to a miniaturization rate equal to 76%. In relation to the initial antenna, the antenna generates a bandwidth of 4.54GHz, from 3.58GHz to 8.12GHz, which means that the antenna covers the following technologies: WLAN Wireless Local Area Network, Radio local area Network, Worldwide Interoperability for Microwave Access Wimax, Hiper LAN2.

Keywords-; patch antenna; genetic algorithm; broadband; miniaturization; defect ground plan DGS.

1 Introduction

The patch antenna is a planar antenna whose radiating element is a generally square conductive surface separated from a conductive reflecting plane by a dielectric plate. Its production resembles a double-sided printed circuit, substrate, and is, therefore, favourable to industrial production. The antenna patch concept appeared in the 1950s, but true development only occurred in the 1970s. It can be used alone or as part of a network. Likewise, it can be integrated as close as possible to the electronic circuits by occupying a reduced volume and conforming to different types of surface. The major disadvantage of this type of antenna is the size which is an area of competition for antenna designers, and the bandwidth likewise. In this paper, we suggest using the DGS (defected ground plan) technique to widen the bandwidth as we use the genetic algorithm to optimize the parameters of the proposed antenna [1-8].

We have structured our article as follows: we will present the geometry of the initial antenna without DGS; then, we will introduce the different forms of DGS in the literature, and we will apply a DGS chosen among these different forms afterwards, and to miniaturize our structure we use the paradigm genetic algorithm, that is why we devoted a part to introduce this algorithm. We end our study with the simulations outcome.

DOI: 10.14738/tmlai.54.3184
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3184

2 Initial Geometry of Antenna

2.1 Patch antenna without defect ground structure

The initial antenna proposed is shown in Figure 1. It is a simple structure with a partial ground plane; the antenna is printed on a Roger FR4 substrate of permittivity 4.4 with a loss tangent 0.02, the length is 50 mm, the width is 40 mm and the height is 1.6 mm. Using microstrip line, we can give excitation to the antenna as shown in Figure1.

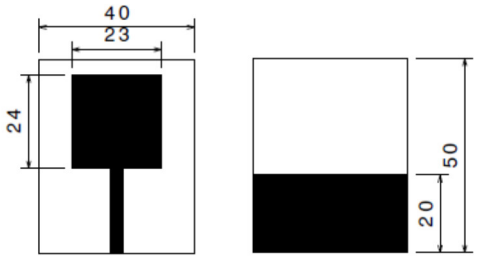


Figure 1. Initial patch antenna "A" without DGS.

The electromagnetic simulation software CST was used to study the performance of our structure. We used this simulator to assess the different radiation characteristics of our patch antenna, particularly the two parameters reflection coefficient and bandwidth.

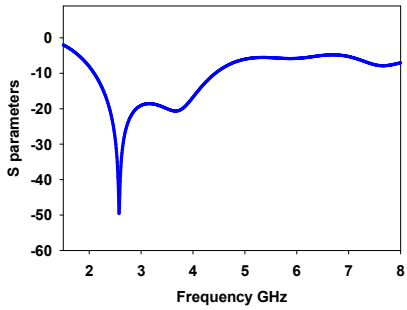
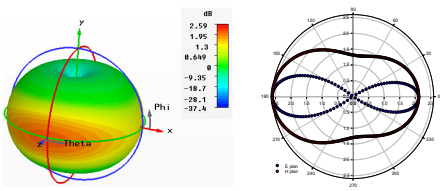


Figure 2 Reflection coefficient S11 paremeters using CST without DGS

It is noted that the initially proposed antenna radiates in the band 2.04GHz- 4.491GHz, with good adaptation; the bandwidth covers the following technologies: ISM BANDS (2.4GHz-2.5GHZ), WLAN Wireless Local Area Network (2.4GHz- 2.84GHz), BLUETOOTH (2.402GHz and 2.480 GHz), or (2.4GHz and 2.483GHz) and almost the entire S-band 2.04GHz-4GHz.



(a) (b)

Figure 2 Simulation results of initial antenna 'A' without DGS : (a) 3D radiation pattern at 2.5 GHZ and (b) 2D radiation pattern at 2.5GHz patch antenna with defect ground structure

The DGS structures are an evolution of electromagnetic bandgap structures (EBG: Electromagnetic BandGap) [5-6]. These EBG structures have aroused the interest of many researchers thanks to their interesting properties in terms of miniaturization and suppression of surface waves. The DGS structures offer the same advantages as the EBG structures with the difference that the use of one or a few cells is sufficient to ensure the same performances. A DGS unit cell consists of a defect (etching) in the ground plane of a transmission line in microstrip, in a coplanar waveguide or in any structure where a ground plane exists. There exist different forms of DGS such as: Dumbbell-shape, Spiral-shaped, H-shaped, U-shaped, arrowhead dumbbell, concentric ring-shaped, split-ring resonators, inter-digital, cross-shaped, circular head dumbbell, square heads connected with U slots, open loop dumbbell, fractal, half-circle, V-shaped, meander lines, U-head dumbbell, double equilateral U and square slots connected with slot at edge [9].

We have chosen to introduce a double iteration of the DGS structure in the form of T in the ground plan of the patch Figure 3.

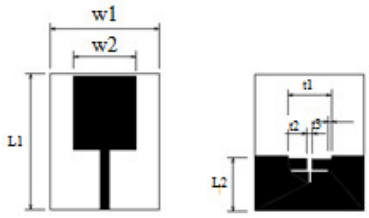


Figure 3 The initial patch antenna with DGS "antenna B"

Table 1 presents the parameters of the patch including the dimension of initial variable od DGS using in the optimization of the prototype (L1, L2, W1, W2, t1, t2, t3)

Table 1 Initial parameter design antenna b 'with DGS'

parameter	*Value (mm)*
L1	50
W1	40
W2	24
T1	16
T2	7
T3	1.2
L2	20

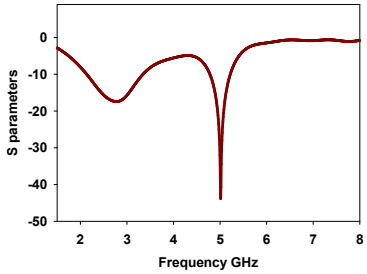
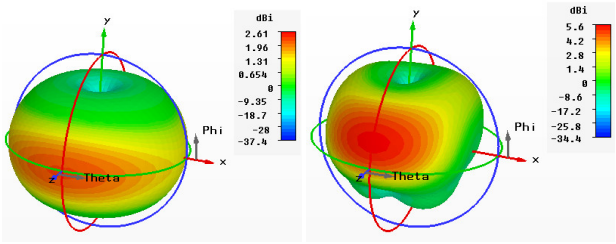


Figure 4 Reflection coefficient S11 paremeters using CST with DGS



(a) (b)

Figure 5 Simulation results of antenna 'B' with DGS : (a) 3D radiation pattern at 2.5 GHZ and (b) 3D radiation pattern at 5GHz

The simulation of return loss with the insertion of DGS shows the generation of another frequency from 4.7GHz to 5.3GHz (WiMAX technology). In the following part, the genetic algorithm is used to optimize the parameters L1, L2, W1, W2, t1, t2 and t3, in order to miniaturize the size of the antenna and widen its bandwidth.

3 Antenna Optimization using Genetic Algorithm

3.1 introduction to genetic algorithm

Genetic algorithms are commonly used in multi-criteria optimizations [10-11]. This paradigm, combined with the terminology of genetics, allows us to exploit genetic algorithms: We find the notions of Population, Individual, Chromosome and Gene.

- The population is the set of possible solutions.
- The individual is a solution.
- Chromosome is a component of the solution.
- The gene is a characteristic, a particularity

The genetic algorithm aims to iteratively select the best sets of parameters that minimize the objective function in 3 steps [11-12].

- Selection: Selection involves selecting the most suitable individuals in order to have a population of solution closest to converging towards the overall optimum,
- Crossing: is the result obtained when two chromosomes share their characteristics. This allows the genetic brewing of the population and the application of the principle of heredity of Darwin's theory. There are two methods of cropping: single or double crossing.
- Mutation: to carry out a random mutation allowing the population evolution.

Figure 6 presents the block diagram of the genetic algorithm [13]

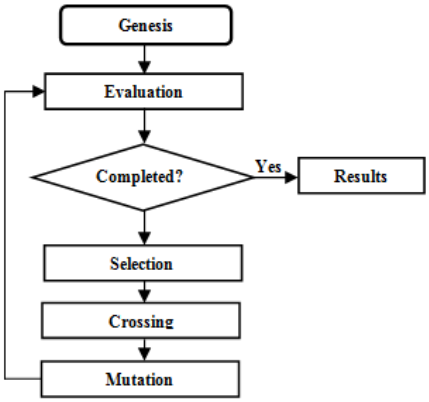


Figure 6 Block diagram of the genetic algorithm

3.2 Optimization procedure : CST-Matlab

We use the MATLAB software to implement the code of the genetic algorithm. The advantage of MATLAB is its power to call external programs. It is a powerful calculator for complex matrix manipulations. For antenna design and analysis, CST is equipped with a VBA script; the genetic algorithm is written under MATLAB calling a VBA script in CST which controls the analysis operations on CST Studio. The important process of the genetic algorithm used to optimize the major parameters of proposed antenna that can be summarized as follows:

- Name the different variables to optimize and define the maximum and minimum limits for each variable to be optimized.
- Population generation, each variable is an optimized chromosome. And write the variables (chromosomes) in line vector forms
- Each variable (each individual of chromosomes), a selection method is applied to select individuals with good characters.
- The selected individuals continue crossing and mutation.
- A new population is generated by continuous reproduction of the offspring. Repetition of evaluation (selection process, crossover, and mutation) until the criterion is satisfied.

The fitness function is defined as follows:

$$\text{fitness: P(x)} = \frac{1}{\text{N}} \sum_{\text{fmin}}^{\text{fmax}} \text{P(f)} \tag{1}$$

$$\text{P(f)} = \begin{cases} S_{11} & for\ S_{11} \geq -10 \\ -10 & for\ S_{11} < -10 \end{cases} \tag{2}$$

S11 is the return loss:

$$S11 = 20LOG \left| \frac{Z_{in} - Z_c}{Z_{in} + Z_c} \right| \tag{3}$$

Where Z_{in} is the input impedance of the antenna and Z_c is the impedance characteristic of the microstrip line.

FIG. 7 shows the flowchart of the genetic algorithm which presents the various steps summarized in two modules: a MATLAB module regrouping the MATLAB code and the other MODULE CST-VBA script

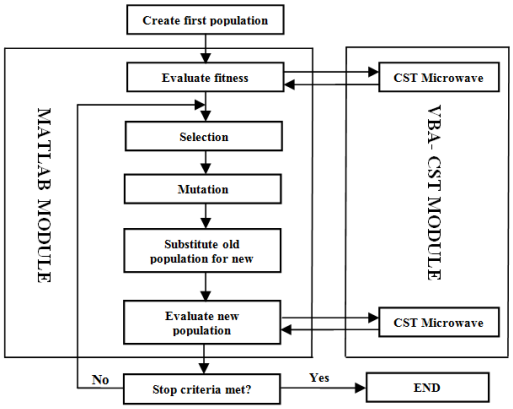


Figure 7 Genetic Algorithm block diagram

Table 2 Initial Parameter Design Antenna B 'With DGS' Optimizing Using Genetic Algorithm

parameter	*Value (mm)*
L1	26.00
W1	20.00
W2	10.00
T1	9.88
T2	6.10
T3	1.12
L2	20
L3 longth of patch	13.09

With these values, the performance of the optimal patch antenna with Defect ground structure is shown in Figure 9.

We can notice that the bandwidth becomes wide and that the optimized antenna covers the band Wimax wholly plus some technologies of the initial antenna. The new large band is "3.58GHz to 8.12GHz".

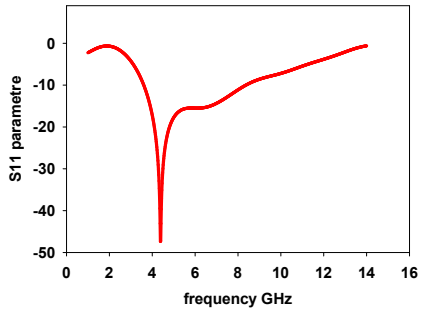
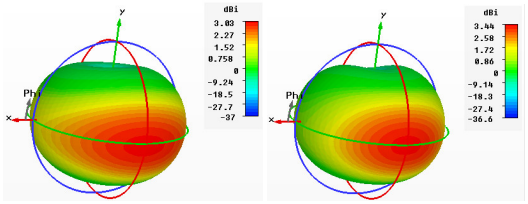
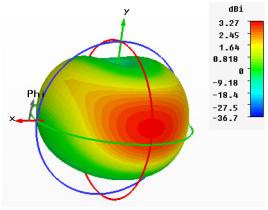


Figure 9 Reflection coefficient S11 with DGS and optimizing with genetic algorithm



(a) (b)



(c)

Figure 10 Simulation results of antenna 'B' with DGS : (a) 3D radiation pattern at 5 GHZ and (b) 3D radiation pattern at 7GHz and (c) 3D radiation pattern at 8GHz

Table 3 Comparative Study

Antennas Table	Characteristics		
	BW GHz	Technology	Size (mm)
Without DGS Antenna 'A'	2.3-4.6	WIFI –Sband	40*50
With DGS Antenna 'B'	2.3-3.7 and 4.7GHz to 5.3GHz	WIFI–Sband-Wimax	40*50
DGS-GA Antenna 'C'	3.58 -8.12	Wide band	20*26

4 Conclusion

The genetic algorithm is used in this work to optimize the parameters of our prototype. The best results have been obtained, that is the miniaturization of the size of the initial patch antenna with a miniaturization rate equal to 76% and the widening of the bandwidth of the antenna to cover other technologies.

REFERENCES

[1] Ayoub, A. F. A., "Analysis of rectangular microstrip antennas with air substrates," Journal of Electromagnetic Waves and Applications, Vol. 17, No. 12, 1755–1766, 2003.

[2] Wong, K. L. and W. H. Hsu, "Abroad-band rectangular patch antenna with a pair of wide slits," IEEE Trans. Antennas Propagat., Vol. 49, No. 9, 1345–1347, September 2001.

[3] Yu, A. and X. X. Zhang, "A method to enhance the bandwidth of microstrip antennas using a modified E-shaped patch," Proceedings of Radio and Wireless Conference, 261–264, Aug. 10– 13, 2003.

[4] Garg, B., R. Tiwari, A. Kumar, and S. K. Thakur, "Design of broadband rectangular microstrippatch antenna inset 'L' shaped feed with rectangular 'L' slots in ground plane," International Journal of Computer Applications, Vol. 29, No. 1, Sep. 2011, ISSN: 0975-8887.

[5] Xiao, J.-K. and W.-J. Zhu, "New bandstop filter using simple defected microstrip structure," Microwave Journal, Sep. 11, 2011.

[6] Ahn, D., J. S. Park, C. S. Kim, J. Kim, Y. Qian, and T. Itoh, "A design of the low-pass filter using the novel microstrip defected ground structure," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 49, 86–93, Jan. 2001.

[7] Omar El Maleky, Farid Ben Abdelouahab, Mohamed Essaaidi, Laila Ajana "Reconfigurable T-Shaped Antenna for S-Band Applications" *Proc. International Conference on Multimedia Computing and Systems Conf.,* October.2016

[8] Omar El Maleky, F. B. Abdelouahab, "Frequency Reconfigurable Patch Antenna for Wireless Communication", *International journal of microwave and optical technology,* Vol. 12, No.1, pp.45-51, January. 2017.

[9] Sung, Y. J., C. S. Ahn, and Y.-S. Kim, "Size reduction and harmonic suppression of rat-race hybrid coupler using defected ground structure," IEEE Microwave and Wireless Components Letters, Vol. 14, No. 1, 7–9, Jan. 2004.

[10] Y. Rahmat-Samii and E. Michielssen, Electromagnetic Optimization by Genetic Algorithms, Y. Rahmat-Samii and E. Michielssen, Eds. New York: Wiley, 1999.

[11] K. M. Luk, R. Chair, and K. F. Lee, "Small rectangular patch antenna," Electron. Lett., vol. 34, pp. 2366–2367, Dec. 1998.

[12] N. Herscovici, M. F. Osorio, and C. Peixeiro, "Minimization of a rectangular patch using genetic algorithms," in Proc. IEEE Antennas and Propagation Int. Symp., vol. 4, Boston, MA, July 2001, pp. 34–37.

[13] H. Choo, A. Hutani, L. C. Trintinalia, and H. Ling, "Shapeoptimization of broadband microstrip antennas using genetic algorithm," Electron. Lett., vol. 36, pp. 2057–2058, Dec. 2000.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Automatic Tool Selection System based on STEP_NC and ISO13399 Standards

Hanae Zarkti[1], Ahmed Rechia[1] and Abdelilah El Mesbahi[1]
[1]Department of Mechanical Engineering, Faculty of Sciences and Technics, Tangier, Morocco.
zarktihanae@gmail.com, rechia@hotmail.com, elmesbahi_abdelilah@hotmail.com

ABSTRACT

Computer Aided Process Planning (CAPP) is considered as an essential component of Computer Integrated manufacturing (CIM) environment and it is an important interface between Computer Aided Design (CAD) and Computer Aided Manufacturing (CAM). The purpose of the CAPP is to determine automatically the use of available resources, including machines, cutting inserts, holders, appropriate machining parameters such as cutting speed, feed rate, depth of cut, and generates automatic sequences of operations and instructions to convert a raw material into a required product, with good surface finish.

The contribution of this work in CAPP field is the development of an automatic tool selection system based on STEP_NC and ISO 13399 standards for turning and milling processes. The paper presents the result of study and analysis of principal system functionalities to be considered. The system consist of four principal modules: a Tool database, feature recognition module, cutting tool selection module and a process optimization module. Finally, based on functional analysis results, the paper present the development of tool database and data tool extraction module from ISO 13399 File using oriented object programing.

Keywords- CAPP; cutting tool selection; functional analysis; database; ISO13399; objected oriented programming;

1 Introduction

The ultimate goal of the factory of the future is to interconnect every step of the manufacturing process.

Today, fast paced technology, sustainability, optimization and the need to meet customer demands have encouraged the transformation of the manufacturing industry, to become adaptive, fully connected and to adapt new standards that involve technical integration of systems across manufacturing processes.

One important stage when such an integration is required is the process planning. Actually, Computer Aided Process Planning (CAPP) represent an important interface between Computer Aided Design (CAD) and Computer Aided Manufacturing (CAM) and its integration is vital to the competitiveness of manufacturers.

Furthermore, the production cost of a manufacturing component depends upon cost of work-piece material, tooling cost, and recurring expenses. Thus, tool selection is considered as an important stage that contributes in the production cost of the manufacturing component. In the early 1980s, research

DOI: 10.14738/tmlai.54.3186
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3186

work was undertaken in the area of computer-aided manufacturing and process planning systems have been developed to select a tool or a set of tools for a specific operation or a set of operations.

Xiaoping Ren, Zhanqiang Liu and Yi Wan developed a computer-based intelligent system for Automatic tool selection for different tool materials. The procedure of selection goes through several main steps: feature specification, machining type selection, cutting tool selection and optimum cutting conditions [1].

Jitender K. Rai, Daniel Brand, Mohammed Slama and Paul Xirouchakis developed a physical model called GA-MPO (genetic algorithm based milling parameters optimization system) in the multi-tool milling of prismatic parts [2].

C.G.Jensen, W.E.Red, J.Pi introduced automatic tool selection methods for five-axis curvatuve matched machining, based on cutter radius, cutter corner radius and cutter effective length [3].

Carpenter and Maropoulos [4] developed a flexible tool selection decision support system for milling operations, the system is called OPTIMUM (Optimized Planning of Tooling and Intelligent Machinability evaluation for Milling).It combine a knowledge based logic and statistical methods.

The authors implemented a module of machinability assessment that produces initial cutting conditions for a wide range of materials and tools. The system also proposes a new optimization criterion related to initial average chip thickness called harshness.

Roshan [5] has developed a graph based algorithm to find the optimal sequence for machining an entire setup that contains a set of features with precedence constraints.

The author presented four extensions to the basic algorithm:

feature level optimization, composite graph method, constrained graph method and sub-graph method.

In the first method, tool sequence graphs are built individually for each feature. In the second method, a composite tool sequence graph is generated for all features in a sibling level.

The constrained graph and subgraph approaches have been developed for situation where different features in the setup have distinct critical tools.

Mookhrejee and Bhattacharyya [6] developed an expert system Extool which automatically selects the turning tool/insert or milling insert, the material and geometry.

The machining time calculation function of the program requires the feed, cutting speed, depth of cut, length of job and the initial and required finished diameter of the job as input.

The outputs are cutting time and number of passes of the tool.

Edalew, Abdalla and Nash [7] developed a system for the selection of cutting tools. It's a dynamic programming based system that utilizes mathematical modules and heuristic data to determine and calculate cutting parameters and total component cost. The system contains the following modules:

the knowledge acquisition module, the knowledge base module, the inference engine, the user interface and the database.

Arezoo, Ridgway and Al-Ahmari [8] developed a knowledge based system for selection of cutting tools and conditions of turning operations. It contains an inference engine, a user interface and explanation facility, a knowledge base and an optimization module for machining conditions.

Although recently, CAPP filed has received more attention from researchers and there have been considerable efforts in developing cutting tool selection systems there still a need for those systems to suit emerging standards of information exchange.

This is the need which the starting point of our research project stems from. Indeed, in the perspective of previous work on CAPP systems based on STEP_NC [9], [10], [11], [12], we aim to develop an automated optimized cutting tool selection system, integrated in a STEP_NC manufacturing chain, capable of assigning adequate tool to manufacture the part automatically and in an optimal manner.

The system comprises four modules: data tool extraction, feature recognition, Tool selection, and optimization modules.

The present work deals with development of data tool extraction module and the tool database.

2 Functional Analysis

We use functional analysis methodology to identify needs and to provide a clear representation of different functionalities of the system proposed.

2.1 Identification of needs and System Functions

The purpose of this design activity is to describe the functions of the system and its parts and indicates the mutual relations. Defining research needs in Fig.1 was the first stage of our work. We used “bêtes à corne” diagram to identify the need of the system. Need answers to three questions: 1 - who or what service makes the system? 2 - on what or who acts on the system? 3 - for what purpose, to do what? .

The next step was to delimit system boundaries by defining existent interactions between the system and different features in its external environment. We used “Pieuvre” diagram to represent these interactions as two sort of functions: Principal functions (PF) and constraint functions (CF). Fig.2 represents connections created between the Tool selection System and the elements of its external environment, considering its using context and each stage of its lifecycle.

2.2 Function decomposition and proposition of solutions

After identification of functions, we develop a graphical representation FAST (Functional Analysis System Technique) Diagram showing the logical relationships between the functions of the system (Fig.3).

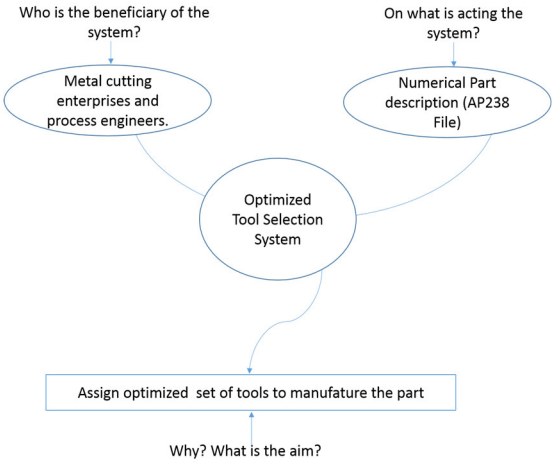
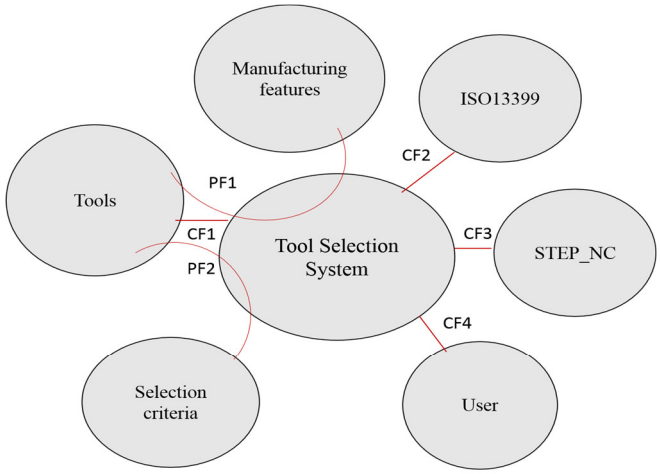


Figure.1. Identification of system need

The FAST diagram aims functions breakdown in subfunctions and eventually components breakdown which is governed by the following intuitive logic type questions like How? From left to right and why? Following the reverse, from right to left. It is used both to represent a known system as well to represent a new product based on its function.



PF1: Assign adequate tools to manufacturing features.
PF2: Optimize Tool selection based on defined criteria
CF1: Automate management of Tools .
CF2: Extract technical data from ISO 13399 file.
CF3: Meet STEP NC standard.
CF4: Provide interaction with user.

Figure.2. "pieuvre" diagram of the system

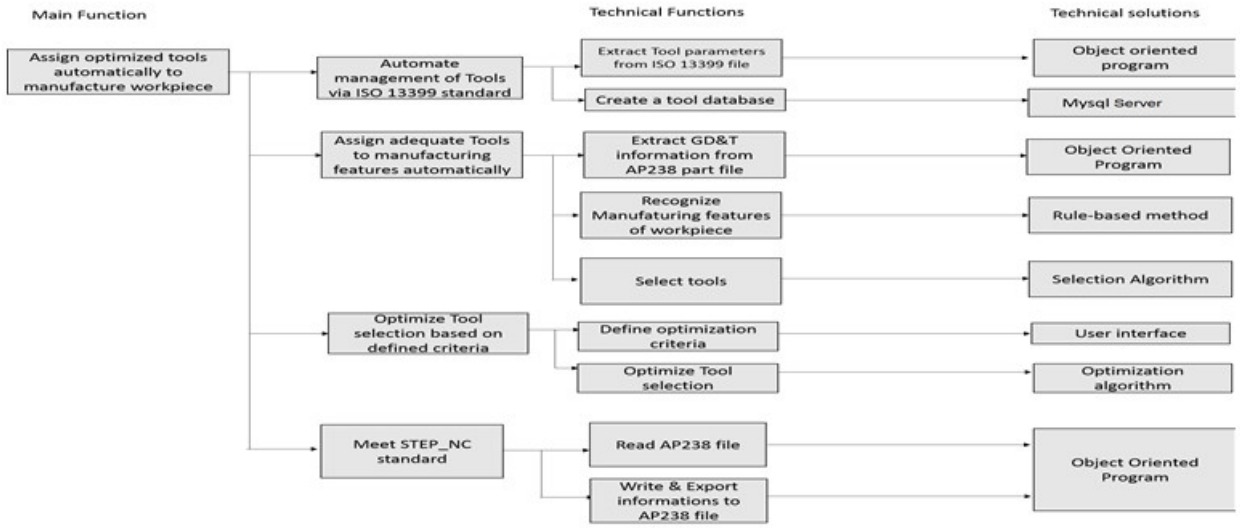


Figure .3 Functional Analysis System Technique diagram

3 Automatic tool selection

3.1 Study of manual tool selection

According to the manufacturer Sandvik Coromant [10], selection of cutters is done manually and goes through five principal stages:

1) Define the operation: The first is to identify the type of operation, then select the most suitable tool considering productivity, reliability and quality.
2) Define the material: Secondly material of the workpiece is defined according to ISO; Steel (P) – Stainless steel (M) – Cast iron (K) – Aluminum (N) – Heat resistant and titanium alloy (S) –Hardened material (H).
3) Select the cutter concept and the cutter: Assess which concept is the most productive for the application, select the cutter and choose the type of mounting.
4) Select the insert: by choosing the insert geometry for the operation and selecting insert grade for optimum productivity.
5) Define the start values: Cutting speeds and feeds for different materials are given on the insert dispensers and in the tables. The values should be optimized according to machine and conditions.

In fact, the manual method presents some drawbacks; on one hand, it depends on the expertise of the user who must manually select parameters and criteria through the machining handbook for a given application, which is time consuming. On the other hand, the choice of tools provided by this method depends essentially on the operations to be performed and not on features to manufacture.

Thus, the aim of our project is to propose an automatic tool selection system in order to reduce Process planning time by using manufacturing feature approach instead of reasoning on operations.

Actually, we consider the part to manufacture as a group of manufacturing features (MF). Each MF has a number properties related to shape, dimensions, technology, tolerances etc.

3.2 Global System Architecture

Based on the results of Functional analysis stage, we defined the global architecture of the automatic tool selection system.

Fig.4 shows flow chart of tool selection systems with inputs, outputs and relations between its modules: Data tool extraction module, recognition module, Selection module and optimization module.

The data tool extraction module is essential to the development of a solid tool database. Extraction procedure starts by uploading the ISO13399 file which contains information about tool, then these information are extracted and stored in the database.

Considering that the system is based on Step_NC standard, the process of tool selection starts by uploading the AP238 file of the Part. This numerical file contains necessary information to manufacture the part. It also comprises geometrical, topological and technical (material, tolerances...) definitions of the part to manufacture.

All of these technical definitions will be extracted using an object oriented program that allow also recognition of manufacturing features.

Then, based on information stored in Tool database, and using an appropriate tool selection algorithm, cutting tools are sorted and affected to Manufacturing features of the part.

A user interface allows planners to determine either to stop selection process and generate STEP_NC AP238 file for machining or to activate optimization module before generation.

In the second case, the user is invited to select optimization criteria and determine its preferences before starting optimization algorithm. Finally, optimal tools will be assigned to manufacturing features and the result is the numerical file STEP_NC AP238 which contains machining process information.

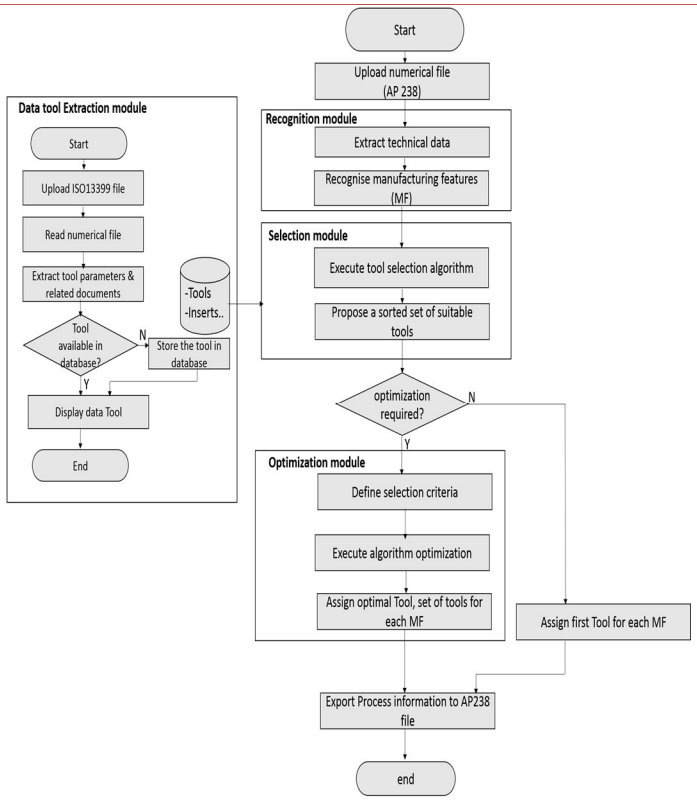


Figure 4. Tool representation in STEP _NC file

4 Cutting tool information exchange

Regarding the poor representation of tool in the older ISO 6983 G and M code language, there is an increasing need for manufacturers to adapt a standard for cutting tool data representation. This urgent need emanates from the fact that the G& M code describes only the path of the tool centre point with respect to machine axes. Consequently, even if the tool selection is done automatically with CAM systems, information about tool (geometry, properties…) will be lost when moving to the machining stage.

ISO 10303 AP238 (namely STEP_NC), and ISO 13399 are emerging standards on which the present work is based. Further information about these standards could be examined in the next section.

4.1 Use of Step-Nc standard

ISO 14649 or STEP-NC is a new standard developed for programming computerized numerical controllers (CNCs). The main development of STEP-NC was carried out in an international IMS/EU project with participants from Germany, France, Switzerland, South Korea, USA, and Sweden. The basic idea is to move toward a description of what to manufacture instead of the current description of how to manufacture. Description of cutting tool data for turning and milling are specified in ISO 14649 part 111[14] and ISO14649 part 121[15] respectively.

ISO 14649 defines a richer model for information transfer between CAD/CAM systems and computerized numerical control (CNC) machines than that of the older ISO 6983 “G and M code” language. In ISO 6983, tools are defined with only their identifiers without any further information (e.g T3), while in a STEP_NC file, tools are defined in express language with values of properties required in machining (e.g cutting

length, hand of tool, depth of cut …) fig 5 is an excerpt of Step_Nc file of example CC2 in annex J in [16] where an end mill tool is defined with its properties.

4.2 Use of ISO13399 Standard

ISO13399 includes the data representation of everything between the workpiece and the machine tool: Information about inserts (e.g. regular and irregular shaped replaceable cutting items), solid tools (e.g. solid drill and solid endmill), assembled tools (e.g. boring bars, indexable drills and indexable milling cutters), adaptors (e.g. milling arbor and chucks), components (e.g. shims, screws and clamps) or any combination of the above can be exchanged.

```
Application object: ENDMILL (#580)
* ITS_ID: #580 ['1']
* EFFECTIVE_CUTTING_DIAMETER: #580, #582, #583, #584, #585
* MAXIMUM_DEPTH_OF_CUT: #580, #582, #583, #584, #586
* HAND_OF_CUT: #580, #582, #583, #584, #587 ['right']
* EDGE_RADIUS: #580, #582, #583, #584, #588
*/
#580=MACHINING_TOOL('1','endmill',(#490),#581);
#581=ACTION_RESOURCE_TYPE('milling cutting tool');
#582=RESOURCE_PROPERTY('tool body','',#580);
#583=RESOURCE_PROPERTY_REPRESENTATION('','',#582,#584);
#584=MACHINING_TOOL_BODY_REPRESENTATION('endmill',(#585,#586,#587,#588),#41);
#585=(
LENGTH_MEASURE_WITH_UNIT()
MEASURE_REPRESENTATION_ITEM()
MEASURE_WITH_UNIT(LENGTH_MEASURE(20.),#554)
REPRESENTATION_ITEM('effective cutting diameter')
);
#586=(
LENGTH_MEASURE_WITH_UNIT()
MEASURE_REPRESENTATION_ITEM()
MEASURE_WITH_UNIT(LENGTH_MEASURE(20.),#554)
REPRESENTATION_ITEM('maximum depth of cut')
);
#587=DESCRIPTIVE_REPRESENTATION_ITEM('hand of cut','right');
```

Figure 5. Tool representation in STEP _NC file

The cutting tool data described include, but are not limited to:

geometrical and dimensional data,

identification and designation data,

cutting material data,

and component connectivity.

The purposes of ISO 13399 and ISO 14649, in regards to tool information, are different. ISO 13399 is a standard for the description of cutting tool as products, while ISO 14649 only describes some simple tool requirements to be used by the CNC when deciding what cutting tool to use [17].

5 Design of database

5.1 Study of cutting tool parameters and data collection

One important step in designing Tool database is to study cutting parameters which are of great importance in selecting cutting tools. Our main focus was on milling process. The result of study [18] shows three types of factors that influence tool selection:

5.1.1 Parameters related to workpiece:

Feature type: The type of feature is considered as a decisive parameter when selecting cutters. Suitable tools are selected by feature category. Indeed, machining features can be divided into two categories,

simple features and complex features. Simple features are realized by using one machining operation, for example, a face and a shoulder can be performed by face milling and shoulder milling respectively. Complex features require at least two machining operations such as curved surfaces and freeform surfaces (features), which require roughing and profiling operations.

Surface roughness: Surface roughness is considered as a critical parameter to select tools and number of applications. For example, a feature having surface roughness of 0.8 (Ra =0.8) have to be machined in three applications, roughing, semi-finishing, and finishing application.

Material of the workpiece: Material of workpiece is a determinant factor for selecting the type of the cutter, the insert grade and the machining parameters. The manufacturer gives the designations of materials and their characteristics such as Specific cutting force and Brinell hardness. Depending on the insert grade and the type of material, the manufacturer gives the maximum chip thickness and the recommended cutting speed.

Dimensions of the feature: Dimensions of the feature represent an important parameter that lead the choice of milling tool. In fact, the depth of the feature is critical to determine the type of the cutter and the insert size, on the other hand the length of the feature represents an indicator of which diameter to choose.

5.1.2 Parameters related to Cutting tool:

Depth of cut (ap): In mm (axial) is what the tool removes in metal on the face from the workpiece. This is the distance the tool is set below the un-machined surface.

Tool material: Different machining applications require different cutting tool materials. The cutting tool material affect directly the cutting hardness, temperature stability and tool life.

Cutting width (ae): in mm (radial) is the width of the component engaged in cut by the diameter of the cutter. It is distance across the surface being machined or, if the tool diameter is smaller, that covered by the tool.

Entering angle: As regards cutting geometry in milling, the entering angle (κr), or the major cutting edge angle, of the cutter is the dominant factor affecting the cutting force direction and chip thickness. The choice of insert geometry has been simplified into three practical areas of varying cutting action effects: Light (L), general purpose (M) and tough (H) geometries.

Zn: The number of available cutter teeth in the tool varies considerably and is used to determine the table feed while the effective number of teeth (zc) is the number of effective teeth. The material, width of component, stability, power, surface finish influence how many teeth are suitable.

5.1.3 Parameters related to machining process

Cutting speed (vc): in m/min indicates the surface speed at which the cutting edge machines the workpiece. This is a tool oriented value and part of the cutting data which ensures that the operation is carried out efficiently and within the recommended scope of the tool material.

Spindle speed (n): in rpm is the number of revolutions the milling tool on the spindle makes per minute. This is a machine oriented value which is calculated from the recommended cutting speed value for an operation.

Feed per minute (f): also known as the table feed, machine feed or feed speed, in mm/min is the feed of the tool in relation to the workpiece in distance per time-unit related to feed per tooth and number of teeth in the cutter.

Maximum chip thickness (hex): in mm is the most important limitation indicator for a tool, for an actual operation. A cutting edge on a milling cutter has been designed and tested to have a recommended starting value and a minimum and maximum value.

5.2 Data dictionnary and Elaboration of relational diagram

The second step after definition of cutting parameters was to classify these parameters as properties of database entities or tables and to determine relationships between these entities via entity relationship modeling. The result is relational diagram of database depicted in figure 6 that we developed under MySQL workbench software. Inheritance relationships between classes were conceived based on tool and cutting item classification in ISO13399 part 2 [19] and ISO 13399 part 3[20] respectively. For example, a facemill derived from Mill class which derived from tool item class.

5.3 Database implementation

We choose to feed database automatically with data from The Sandvik Coromant [13] ISO 13399 catalogue for milling and turning tools.

For this purpose an object oriented program was created using c# programming to extract technical data from ISO file and ensure data feed and update of database. More details on program functionalities are given in the following section.

6 Extraction of Cutting Tool Data

6.1 ISO 13399 file structure

An ISO 13399 exchange file contains an electronic representation of cutting tool data as defined by the information structure that can be exchanged by tooling applications.

ISO 13399 is defined as a Schema in the ISO 10303 Express language. The ISO 13399 schema develops a framework in which to define tooling properties, tooling assemblies, and relationship between tooling elements. There are no actual tooling properties in the ISO 13399 schema. Instead ISO 13399 develops a programming structure in which to embed ISO 13584, which is the Industrial automation systems and integration – Parts library (PLIB)[21].

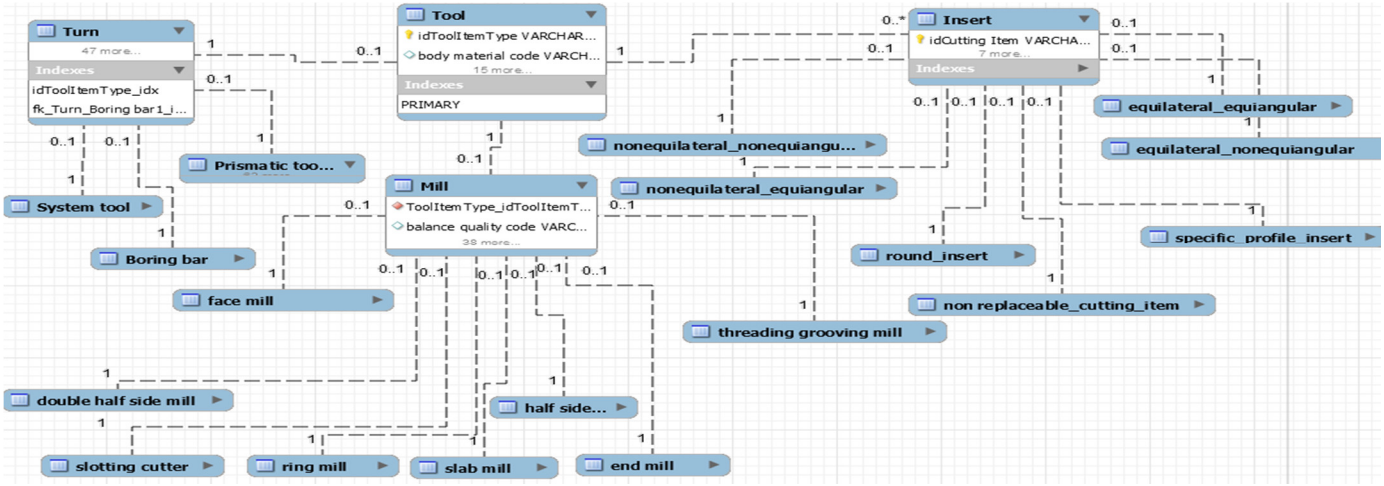


Figure 6 Database relational diagram

Figure 7 illustrate the main entities and relationship between them.

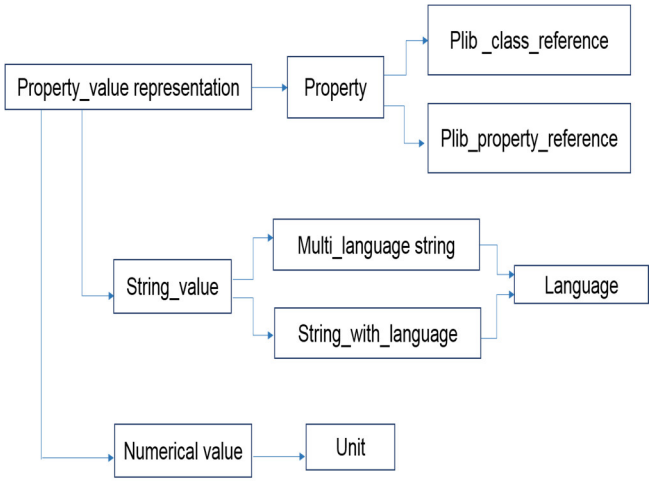


Figure 7 Sample Property Value Information Model Structure

A property_value_representation is used to define properties, based on a property and a value. Each property has a PLIB class and property reference. Definitions of entities linked to property_value representation class are as follow:

Property: A property is the definition of a particular quality. A property may reflect physics or arbitrary user defined measurements.

Plib_class_reference: A plib_class_reference designates a class in a library compliant to ISO 13584 (Parts Library).

Plib_property_reference: A plib_property_reference designates a property in a library compliant to ISO 13584.

String_value: A string_value represents a sequence of one or more alphanumeric characters. A string_value is a type of property_value.

Multi_language string: A multi_language_string represents text information, expressed in one or more languages, associated with object.

String with language: A string_with_language represents text information in a specific language together with an identification of the language used.

Language: A language is a specification of the language in which information is given.

Numerical value: The specified_value specifies the property_value that qualifies the property_value_representation by a value_with_unit,.

Unit: specifies the unit in case the property value is with unit.

6.2 Extraction of cutting tool data

In this section we present the object oriented program that we developed using c# programming. The extraction procedure and results are shown in figure 8.The purpose is to provide a fully integrated system that enables feed and modification of cutting tool database with recent tool information.

Tool data are extracted automatically from the ISO13399 file and stored in the database. After examination of the file data model, entities in figure 7 were mapped to c# classes to be used in different part of the program. Extraction data tool scenario is as follow:

The user is invited to load the P21 file which is displayed in figure 8 (1). after a simple click, the extraction program is run and tool information are given in details in (2) the example shows a face milling cutter 'CoroMill 357' with different parameters. Also, names of correspondent inserts are displayed in a list box (3), and details could be reached after button click to run insert parameters extraction program and display results in (4) .Finally, the user could add the new item to database or update information in (5).

7 Conclusion

The aim of this paper was to dress the architecture of our automatic optimized cutting tool selection system which is currently under development, based on functional analysis methodology. The paper dealt also with the development of the data tool extraction module and database that will be of great importance to the automation of cutting tool selection.

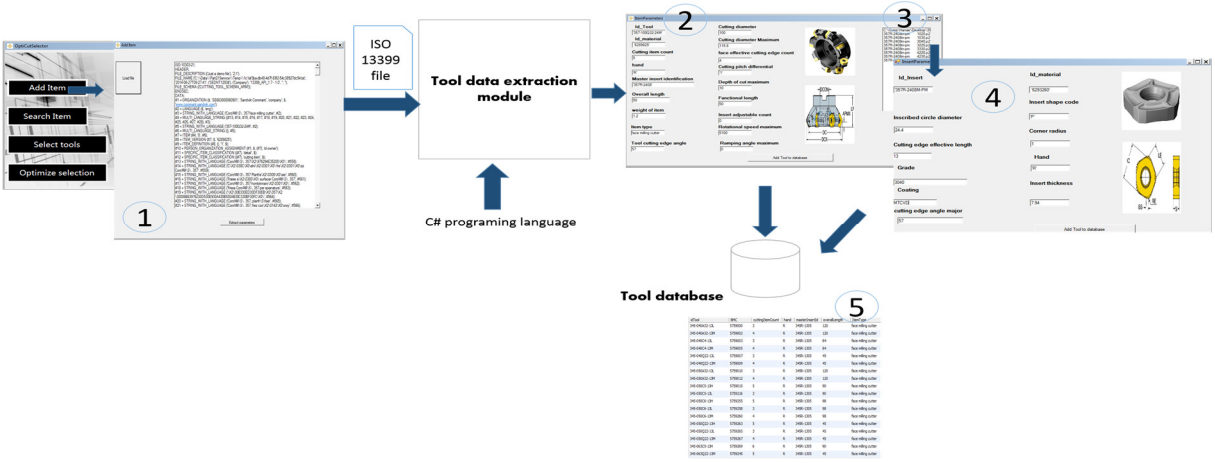


Figure 8. Cutting tool data extraction module

The data tool extraction module was the automatic solution to feed and update database. Since tools manufacturers tends to adopt ISO13399 standard, we consider to update database with recent tool data to provide an effective way of tool selection. Development of feature recognition, tool selection, and optimization modules will be the object of future works.

REFERENCES

[1] X. P. Ren, Z. Q. Liu, Y. Wan, "A Computer-Based Intelligent System for Automatic Tool Selection", Materials Science Forum, Vol. 723, pp. 238-242, 2012.

[2] Jitender K. Rai , Daniel Brand , Mohammed Slama & Paul Xirouchakis (2011) Optimal selection of cutting parameters in multi-tool milling operations using a genetic algorithm, International Journal of Production Research, 49:10, 3045-3068, DOI: 10.1080/00207540903382873.

[3] C. G. Jensen, W. E. Red, and J. Pi, "Tool selection for five-axis curvature matched machining," Comput. Aided D., vol. 34, pp. 251-266, Mar. 2002.

[4] I.D.Carpenter,P.G.Maropoulos, A flexible tool selection decision support system for milling operations, Journal of Materials Processing technology 107 (2000) 143-152.

[5] Roshan M.D'Souza, On setup level tool sequence selection for 2.5-D pocket machining, Robotics and Computer Integrated Manufacturing 22 (2006) 256-266.

[6] R.Mookherejee, B.Bhattacharyya, Development of an expert system for turning and rotating tool selection in a dynamic environment, Journal of Materials Processing Technology 113 (2001) 306-311. Figure 5.Tool data extraction module

[7] K.O. Edalew, H.S. Abdalla, R.J.Nash, A computer- based intelligent system form automatic tool selection, Materials and design 22 (2001) 337-351.

[8] B. Arezoo, K.Ridgway, A.M.A. Al-Ahmari, Selection of cutting tools and conditions of machining operations using an expert system, Computers in industry 42 (2000) 43-58.G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

[9] Abdelilah Elmesbahi, Ahmed Rechia, and Oussama Jaider, "optimized automated choice of cutting tool machining manufacturing features in milling process", proceeding of the 11th World Congress on Computational Mechanics (WCCM XI), Barcelone, 2014.

[10] Hanae Zarkti, Abdelilah El Mesbahi, Ahmed Rechia, Oussama Jaider. Towards an Automaticoptimized tool selection for milling process, based on data from Sandvik Coromant. X`eme Conf´erence Internationale : Conception et Production Intégrées, Dec 2015, Tanger, Morocco.

[11] Oussama Jaider, Abdelilah El Mesbahi, Ahmed Rechia, Hanae Zarkti. An automatic Feature-based tool selection approach for turning process based on data from Sandvik Coromant. Xème Conférence Internationale: Conception et Production Intégrées, Dec 2015, Tanger, Morocco.

[12] Oussama JAIDER, "Planification automatique des processus d'usinage des pièces de révolution, basée sur les normes STEP et STEP-NC", Thèse de doctorat, Faculté des Sciences et Techniques, Tanger, Maroc, 2017.

[13] www.sandvik.coromant.com/

[14] ISO 14649-111: Industrial automation systems and integration — Physical device control — Data model for computerized numerical controllers — Part 111: Tools for milling machines.

[15] ISO 14649-121: Industrial automation systems and integration — Physical device control —Data model for computerized numerical controllers — Part 121: Tools for turning machines.

[16] ISO 10303-238: Industrial automation systems and integration — Product data representation and exchange — Part 238: Application protocol: Application interpreted model for computerized numerical controllers.

[17] OLOF NYQVIST, Information Management for CuttingTools. Information Models and Ontologies, Doctoral thesis, Stockholm, Sweden 2008.

[18] Graham T. Smith, cutting tool technology: industrial handbook, Springer, 2008.

[19] ISO/TS 13399-2: Cutting tool data representation and exchange — Part 2: Reference dictionary for the cutting items.

[20] ISO/TS 13399-3: Cutting tool data representation and exchange — Part 3: Reference dictionary for the tool items.

[21] ISO/TC184/SC4 (2001a). Industrial Automation Systems and Integration – Parts Library – Part 1: Overview and fundamental principles. International Organization for Standardization. ISO 13584-1.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Remote-controlled Laboratories of Experimental Physics: Measuring the Stiffness of a Spring

AL Sabri Ahmed1, Khazri Yassine[1], Moussetad Mohamed[1], Akensous Youness[2], Fahli Ahmed[3]
[1]*LIMAT, Faculty of Sciences Ben M'sik, Hassan II University Casablanca, Morocco*
[2]*GAGE, Faculty of Sciences Ben M'sik, Hassan II University Casablanca, Morocco*
[3]*Faculty of Sciences and Techniques, Hassan I University Settat, Morocco*
a.alsabri2011@gmail.com

ABSTRACT

In modern E-Learning, students and scientists will be able to access Web portals for scientific computer and data infrastructures, thus accessing large collections of data and digital objects using metadata, knowledge management techniques, and specific data services. Students will apply existing scalable Web and grid technologies to access and share scientific data, using educational and computing resources to run scientific Practical exercises. Such an approach will allow the creation of enriched interactive through a real devices and real remote mechanisms that interactively support the exploration of scientific phenomena. Advanced repository and collaboration services will allow students to remotely and securely up- and download science and engineering learning objects. E-lab is essentially created to realize experiments by interacting with real devices that are real remote mechanisms, through an appropriate telecommunications platform, equipped with a dedicated management system and a number of software interfaces and material. Our works is aiming to: Measurement of the rigidity of a spring. This method translator pedagogical experiences and turn them into reality at an affordable, expand research sources to students.

Keywords-learning; data infrastructures; scientific computer; data services Practical exercises; E-lab; experiences experiences.

1 Introduction

Since 1990s, there have been consistent efforts in developing web-based education environment at many institutions around the world. Many educators have strong feelings about the relative merits of different technologies to be used in engineering laboratories. The argument is significant because it is clear that the choice of laboratory technologies could change the economics of engineering education, and it is also clear that changing the technology could change the effectiveness of education [1].

The Information and communication technologies (ICT) play key roles in transforming traditional laboratory into flexible and open learning spaces. It includes any communication device or application, encompassing: radio, television, cellular phones, computer and network hardware and software, satellite systems and so on, as well as the various services and applications associated with them, such as videoconferencing and distance learning. ICTs are often spoken of in a particular context in education, health care, or libraries. In this paper, we use this technology in education of practical experiences. So the

DOI: 10.14738/tmlai.54.3187
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3187

practical experience is an important component of the educational process. However, the time and economical resources often required for the setting up and construction of scientific laboratories is outside the scope of many institutions [2] [3]. The practical work of laboratories is subject to constraints for example: students with disabilities may have less of a chance than other students to interact in laboratory environments. Thus the remote laboratories make an economic problem to Moroccan universities, also an important facilitator for students who, for any reasons must take courses remotely in order to increasing number of students. Moreover, the advantages and motivations: protecting the students while conducting experiments involving health hazards (e.g., radioactivity, UV …), eliminating time and space constraints through affording flexibility, affording the opportunity for conducting an online experiment with real equipment anytime and anywhere and reducing the cost of lab equipment which is expensive [4] [5].

2 Technology background

2.1 Technology and Architecture

Our previous research was based on the use and construction of remote laboratory now we describe the technology that has been constructed, so that we will provide a context for the discussion of the assessment model for those whom they are not familiar with remote laboratory. The Remote Laboratory Experimentation represents an extension to the ways in which people utilize the Internet. A remote laboratory for engineering education should achieve an integrated environment for the user controlling the actual device in the remote site and conduct the actual experiments in the remote laboratory via a computer network [6]. The core of the Remote Laboratory is a set of general and / or specialized instruments linked to a set of personal computer systems connected to the Internet. With the ability to remotely configure instruments and data analysis through software, the laboratory will facilitate the sharing of expensive instruments and equipments. The development of a remote laboratory includes the analysis of user requirements, remote control functionalities, simultaneous user operation, sharing the online data from an experiment, read data, change variables and controlling equipments. To achieve the above, a distributed client-server environment has been designed (the hardware architecture for the remote laboratory system is shown in Fig. 1. Using this architecture, the learners will be able to send themselves, orders through the web browser. Computers in the Remote Lab are connected to engineering instruments. When students log in to these machines over the Internet, they are able to control both the computer and the equipment. A video camera can also be used to broadcast live, what is happening in the physical world. It does not matter if the student is in a dormitory nearby or on the other side of the world [7].

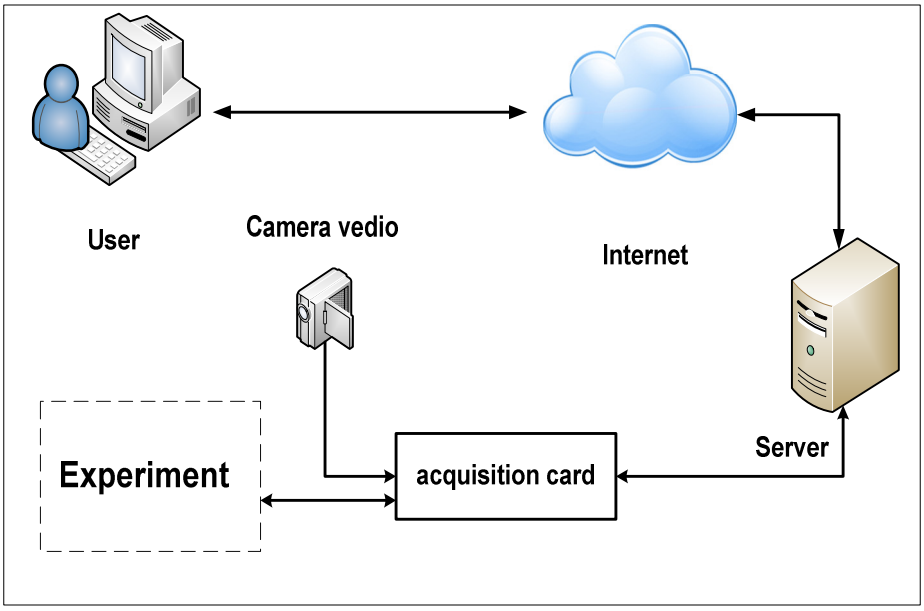


Figure 1 Architecture of Remote Laboratory

2.2 Study Theory of Measuring The Stiffness of a Spring

Mechanic's experience of material point represents an example for the experimental setups implemented so far on our project. An experimental setup for calculating the stiffness of a spring was designed in a modular fashion, which allows straightforward extension to multiple degrees of freedom as is shown in Fig. 3. It is proposed in this manipulation to measure experimentally the stiffness K (spring constant). A spring constant is the measure of the stiffness of a spring. If a mass m is attached to the lower end of the spring with empty length Lo, the spring stretches a distance d from its initial position under the influence of the "load" weight shown schematically in Fig. 2.

$$P = mg \tag{1}$$

The downward force ***P*** must be balanced by the upward restoring force of the spring when the system is at rest.

$$F = P = mg \tag{2}$$

According to Hooke's law the restoring force of the spring is directly proportional to the elongation within the elastic limit (the maximum a spring can stretch without being permanently deformed) and can be written as

$$F = -kd \tag{3}$$

Where F is the restoring force in Newton (N), k is the spring constant in (N/m) and d is the elongation (the stretched amount) in Meters (m). The restoring force is in the opposite direction to its elongation, as shown by the minus sign. When the applied force (the loaded weight) is used in the calculation, the minus sign is not used and the above equation can be written as:

$$\boldsymbol{F = k\ d\ or\ mg = k\ d} \tag{4}$$

The spring constant, k, can therefore be calculated.

$$\boldsymbol{K = \frac{F}{d} = \frac{mg}{d}} \tag{5}$$

Where

$$\boldsymbol{d = L_{eq} - L_0} \tag{6}$$

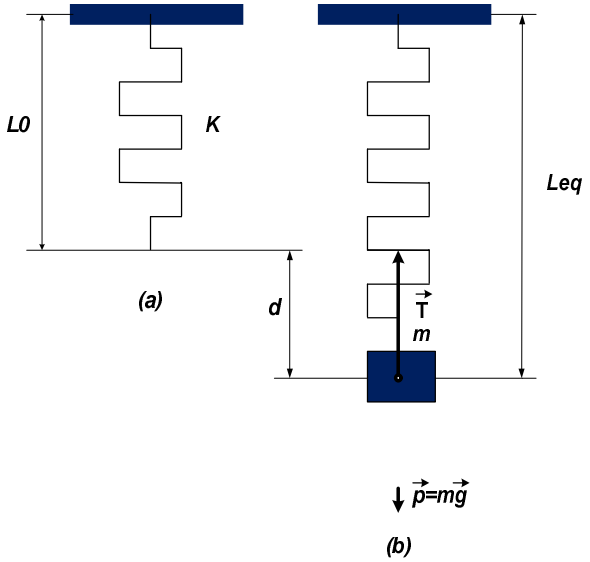


Figure 2 Mechanical system of mass-spring
(a): Spring not loaded, (b): Spring loaded in equilibrium

2.3 Pedagogical use of the Remote Laboratories

The remote laboratory experiments are introduced to students progressively. Here we describe the three phases that the instructor leads the student through by using a video that will be presented in our interface. In the first phase, the students are tasked with re-familiarizing themselves with the underlying physical principles, the experimental equipment and procedure of the particular experiment to be performed. A guidebook (Supporting background materials) is provided also an online documentation [8].

The second phase of the laboratory experience is directed to the laboratory facility under using a tutorial video and a tutorial guide (how to use the technology). This experimental portion is designed in such a fashion that it fits within a time period that is significantly shorter than a traditionally structured laboratory session. Thus this approach retains some – albeit brief by traditional standards – elements of a hands-on experimental experience. Furthermore, the students can watch a video from the instructor describing all steps, thus facilitating the later execution of the remote laboratory assignment. In the third stage, the students continue more detailed experimental studies in a platform of a remote lab (remote fashion way). Typically, one week is given for the student to complete the experiment and filled in the report of the experiment. The described approach leverages a limited number of experimental stations for a wide variety of studies [1] [9] .

2.4 Usage of the Remote Laboratories

The remote laboratory project presented in this paper is illustrated by the following example; the measurement of a spring stiffness. During First step of laboratory usage, the students can run a free test after; connection to the robot is completed. All technical issues related to the use of the robot and calibration of the system as well as data acquisition can be on the platform of the laboratory (represented in a video and documentation). In order to avoid the repetition of experimental procedures, the selection part of the type of experience, facilitates and avoids the students the repetition of inputting data. After

describing how the students will perform the manipulation, the data in the form of measurement carry out by the platform will be elaborated to complete a form and make the necessaries calculus [1] [10].

3 The e-Lab system

In order to define a Remote Laboratory system, enabling real laboratory equipment (like a spring) to be remotely controlled by the student using a domestic Internet connection, we collected the following requirements:

Students should not need any special software to carry out their experience, and then it must be Web-based. As any Web page, it should accessed by means of a standard Web browser. The implementation should take on consideration web protocols and common components to guarantee the desired Web compatibility. It must manage security features

A coordinating supervisor should manage the laboratory session, to ensure security and managing of futures, authorize the control requests and to protect the equipment against any potentially damaging operation.

A standard approach must be defined to cover a large range of computer controlled lab equipments, so that the e-lab will be reusable over different lab types.

The approach should consider the main features of a collaborative Web application.

The e-lab platform scenario is that of a real classroom, which must remotely use a lab equipment located anywhere in a real laboratory. The real classroom is composed of a speaker (video made by a tutor) who will describe and comment the experience, as well as facilitating the use of the platform for the student connected to the Internet at least via a domestic ADSL. Students can remotely use the equipment and they can remotely control it [6] [7].

3.1 The e-Lab Platform

According to the above requirements, we will describe our e-lab platform project, which is a remote control of the robot (will be used to move mass then the Measuring the stiffness of a spring) of our Research Department. The main components of the e-Lab system are:

3.2 Hardware and Software Tools

To realize our system we use the next compensates:

Hardware

Acquisition card (Arduino uno)
Server
4 Servo motors
Camera video
Robotic Arm

Software

Visual basic (Microsoft Visual Studio 2010 Express).
Arduino 1.6.5

3.3 Mechanical system

In order to automate the mass-spring mechanism we modified the system to command our system remotely easily. Our model consists of a spring, ultrasonic sensor, mass and a reflection disk. The ultrasonic sensor is fixed at the top of the system to emit the waves towards the reflection disk (transceiver) and send the data to the acquisition card during the mass change. The Fig. 3 illustrates a mechanical system

3.4 Acquisition Card

The data controlling unit is an embedded system platform. For this propose, we are used an Arduino Uno. It is powered by 9V from power & programmed for the specific applications using Arduino open source software. The robot consists of 4 servo motors to be able to move smoothly into four themes, for moving the masses and hooking them to the determination of the spring. The 4 servo motors are connected to Arduino Uno, by Pin 8, Pin 9, Pin 10 and Pin 11 to send the control of each motor whereas an ultrasonic sensor is connected to Arduino Uno by Pin Analog.

This sensor collects information data and sends it to the Arduino Uno for digitally processing analysis of data. The ultrasonic sensor takes the information data from the mechanical part and send it to acquisition card[8]. The electronic diagram of Acquisition Card is shown in Fig. 4.

3.5 The User Interface

The e-lab user interface design and customized as a web application (shows Fig. 5). The design includes three task areas: on the left-hand side there is the video area, showing what happens in the lab room. The middle side of the screen contains the area with the Tools Panel; the Tools Panel is organized with thumbnails each displaying an experimental test, also use equipment of the lab. The left side of the application contains the field of connection to the lab equipments (robot).

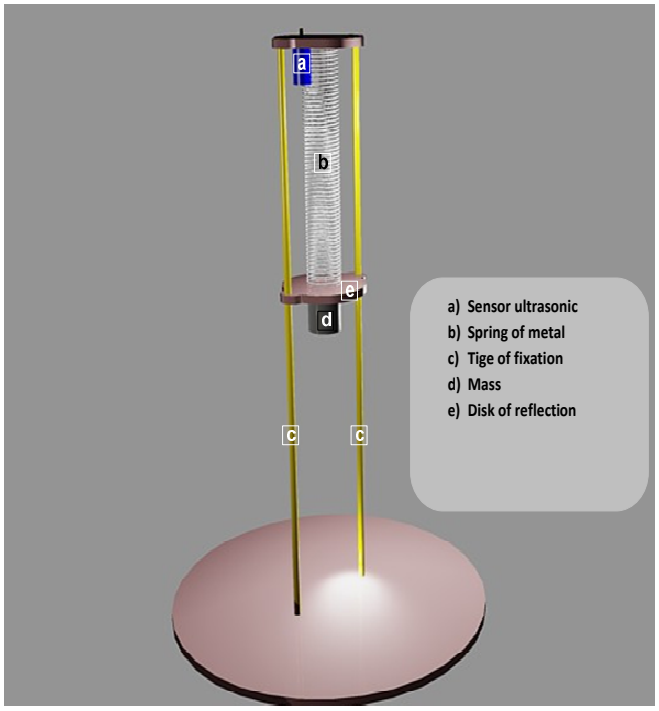


Figure 3 Design model of mechanical system

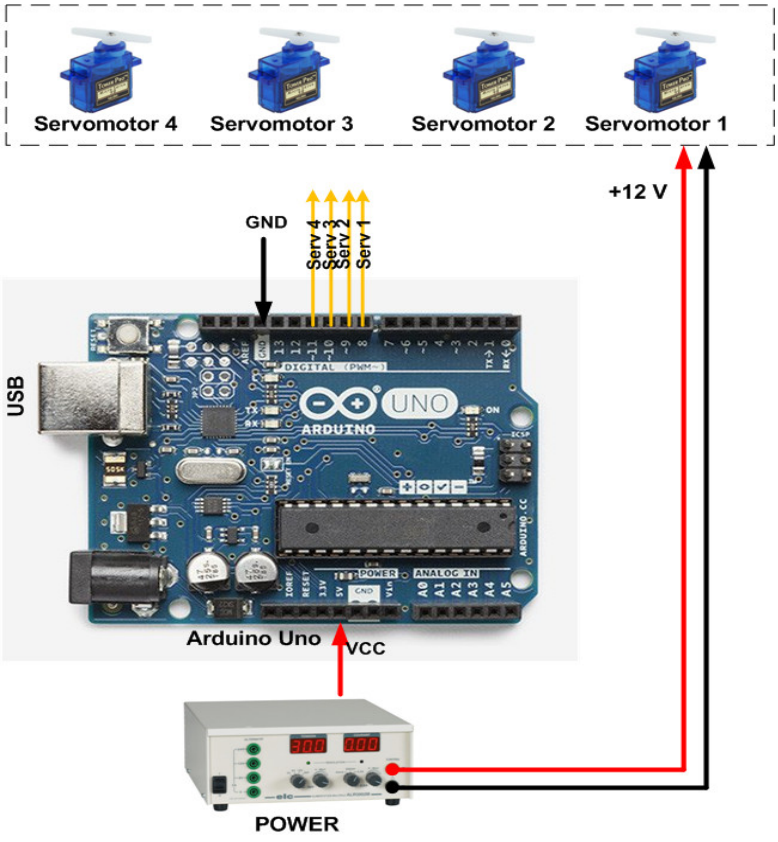


Figure 4 The electronic diagram of Acquisition Card

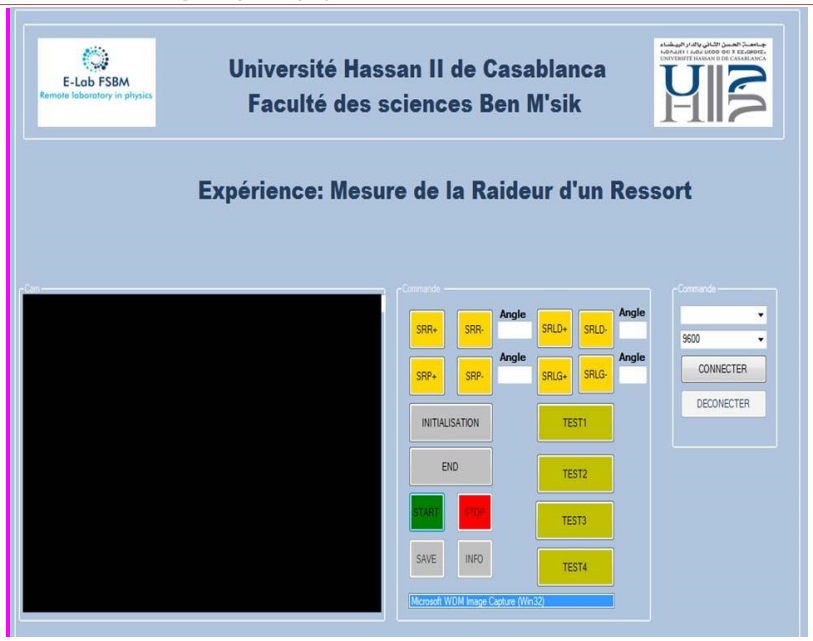


Figure 5 Application Interface

4 Discussion

The paramount importance of the establishment of remote laboratories is to prepare an easy environment, flexible as possible for all students. The e-lab user interface has been designed with the intent to be easily usable for all categories of students, who are accustomed to video games, chats and video conferences. The technique represented in this article aims to reduce the financial and human resource costs for all establishments concerned. As well as any problems of the dangerous kind of experiences and the difficult manipulations will disparate. The Fig. 6 illustrates remote laboratory on a real time.

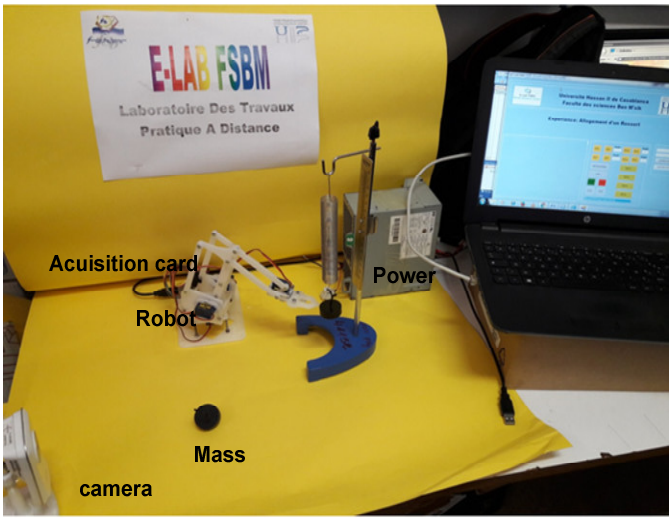


Figure 6 Remote Laboratory in Real Time

5 Conclusion

The forms of Internet learning have been expanded including live physical experiments that much more complicated to implement compared to lecture modules. In this project, we first compared remote laboratories and practical laboratories. The result is a remote laboratory platform based on a web technology, the remote laboratory environment represented in this paper is e-lab platform Constituted by those equipments locally controlled by a computer (robotic arms). As a platform e-lab is evolving into a Reusable service web application, it can control other laboratory equipments then unite them on a unique virtual workbench. For more we are looking in future to create a chatting space because when students become confused, they often attempt to alleviate the confusion by communicating with comrade, or with the professor then the platform will solve this problem, Also give the possibility to tutors to supervise the students work online, as well as they can indicate them on a real time.

REFERENCES

[1] J. V. Nickerson, J. E. Corter, S. K. Esche, and C. Chassapis, “A model for evaluating the effectiveness of remote engineering laboratories and simulations in education,” Comput. Educ., vol. 49, no. 3, pp. 708–725, Nov. 2007.

[2] A. Alexiou, C. Bouras, and E. Giannaka, “Virtual laboratories in education,” in Technology Enhanced Learning, Springer, 2005, pp. 19–28.

[3] O. Dziabenko and J. García-Zubía, IT Innovative practices in secondary schools: Remote experiments, vol. 10. Universidad de Deusto, 2013.

[4] A. A. TALEB, A. FAHLI, and M. MOUSSETAD, “Mise en Oeuvre d’une Télé-Séance de TP de Physique Nucléaire" Rétrodiffusion de Particules α.”

[5] D. MECHTA, “Plate-forme pour les travaux pratiques à distance sur le Web,” Ferhat Abbas, Algrie, 2012.

[6] M. A. Bochicchio and A. Longo, “Hands-On Remote Labs: Collaborative Web Laboratories as a Case Study for IT Engineering Classes,” IEEE Trans. Learn. Technol., vol. 2, no. 4, pp. 320–330, Oct. 2009.

[7] U. Hernandez-Jayo and J. Garcia-Zubia, “Remote measurement and instrumentation laboratory for training in real analog electronic experiments,” Measurement, vol. 82, pp. 123–134, Mar. 2016.

[8] M. R. Laskar, R. Bhattacharjee, M. S. Giri, and P. Bhattacharya, “Weather Forecasting Using Arduino Based Cube-Sat,” Procedia Comput. Sci., vol. 89, pp. 320–323, 2016.

[9] M. Teng, H. Considine, Z. Nedic, and A. Nafalski, “Current and Future Developments in Remote Laboratory NetLab,” Int. J. Online Eng. IJOE, vol. 12, no. 8, p. 4, Aug. 2016.

[10] C. Depover, F. Orivel, and Institut international de planification de l’éducation, Les pays en développement à l’ère de l’e-learning. Paris: Unesco, Institut international de planification de l’éducation, 2012.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

ZCMGenerator: Generation of ZCM Models from ZC2M Metamodel Based on MDA and ADM Approaches

Abdelaziz Mamouni, Abdelaziz Marzak, Abdessamad Belangour, Mohamed Azouazi, Raki Youness, Zayed Al haddad
Faculty of Sciences Ben M'sik, Laboratory of Information Technology and Modeling, Casablanca, Morocco, Hassan II Universty of Casablanca, UH2C
Cdt Driss El Harti, B.P 7955, Sidi Othman, Casablanca-Morocco
mamouni.abdelaziz@gmail.com, marzak@hotmail.com, belangour@gmail.com, azouazii@gmail.com, raki.youness@gmail.com, alhaddadtri@gmail.com

ABSTRACT

Zakat Calculation Models (ZCM) consist of sequences of specific concepts of the Zakat domain such as category, sub-category, wealth and rate. The content of a ZCM model depends on the school of jurisprudence and the type of Zakatable wealth. Correspondingly, creating a ZCM model involves four steps (analyzing, modelling, developing and generating), the last step consists of generating ZCM model in a structured XML format in accordance with our ZC2M (Zakat Calculation Meta-Model) metamodel which we have already proposed in our previous work. This paper proposes ZCMGenerator, a metamodel-based framework for modelling and generating of the ZCM models adaptable to the new technologies and specific to each school of jurisprudence. These models are extensible, reusable, portable, and can be communicated to any platform without taking into account technical specifications. This paper also presents some cases study scenarios emphasizing the benefits enabled by the proposed framework and a detailed comparison between ZCMGenerator and some existing ZCP (Zakat Calculation Platforms) platforms. Experimental results show that, compared with traditional ZCP platforms, ZCMGenerator is more scalable and high-performance.

Keywords-component; ADM; MDA; ZCM; ZC2M; Zakat; MOF; Metamodel

1 Introduction

Zakat calculation platform or ZCP for short is a platform created to support the understanding and calculating of Zakat. These platforms are becoming an important tool and play a crucial role in Muslims life. Over the past few years, several ZCP have been developed due to the high demands on such Islamic platforms. One of the key aims of these platforms is to support Muslims in calculating their Zakat easily, correctly, quickly and accurately. Actually, they have made the entire process of calculating Zakat extremely simple and straightforward. However, this multitude of platforms, although it has advantages, it also presents a number of disadvantages, for instance, the vast majority of them were developed in traditional manner, most of them have lack of portability, reusability and interoperability. Moreover, they were developed to run solely on one technological platform. Furthermore, they are inflexible and difficult to maintain and change. In order to go some way to overcome these problems, the ZCMGenerator

DOI: 10.14738/tmlai.54.3188
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3188

framework has been developed to fulfil the Zakat calculation requirements by facilitating ZCM models modelling and ensuring reuse and portability. To develop and implement this framework, we have already proposed in [1] a new ADM-based process named ZCMGProcess. The architecture of this process is depicted in Figure 1 which is divided into three main phases. The first phase consists in understanding, describing and comparing the ZCP models structure, where the goal is to generate the General PIM. The second phase aims to enrich this PIM model, where the goal is to generate a new improved model. The third phase aims to generate the ZCM models in XML files. These phases are in accordance with the three stages defined by the ADM (Architecture Driven Modernization) approach that are reverse engineering, restructuring, and forward engineering.

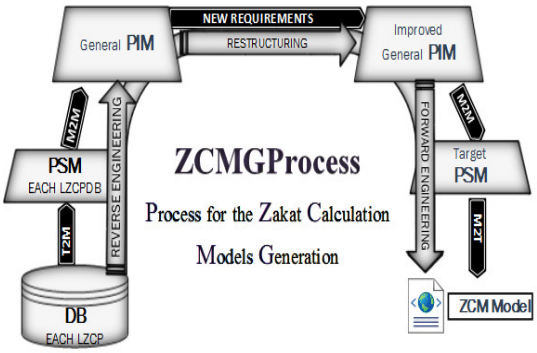


Figure 1. Architecture of ZCMGProcess

The aforementioned ZCMGProcess was inspired on idea of ADM approach. It aims at guiding designers in the modelling of extensible, reusable and portable Zakat calculation models using MDA (Model Driven Architecture) concepts. It not only takes advantage of the MDA approach but also integrates other important aspect of portability, interoperability and reusability that is XML technology. Indeed, the XML brings a number of powerful capabilities to information modelling such as heterogeneity, extensibility, flexibility, internationalization and many others [2]. This process has been defined by bearing in mind its use by developers who wish to design a new ZCP platform and to produce the extensible, reusable and portable models which can be communicated to any platform without taking into account technical specifications.

Then, we have proposed in [3] a new metamodel called ZC2M (Zakat Calculation Meta-Model) (Fig. 2) which is a generic metamodel enabling, among other things, to take into account the specific calculation rules of each school of jurisprudence when determining the Zakat amount. For the sake of providing a comprehensive metamodel, we have used the concepts, which are generally accepted by all schools of jurisprudence. It is given after an in-depth study of the Zakat domain [4] and a survey of the existing ZCP platforms. It is based on the MOF (Meta-Object Facility) and aims to define the key concepts used for modelling the ZCM models and to modernize the existing platforms. It is the first metamodel of its kind to tackle the Zakat calculation problem. Its structure is divided into hierarchical elements viz; Model, Category and Wealth, whose categories can be divided into several sub-categories.

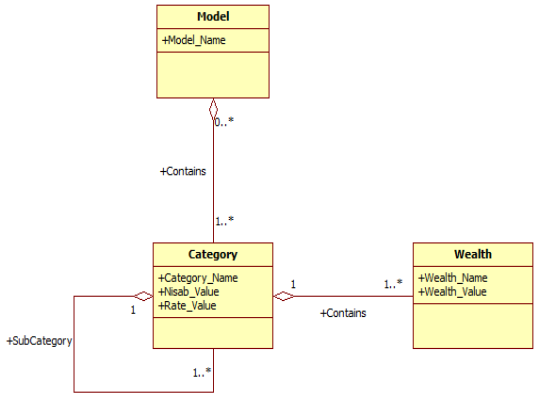


Figure 2. ZC2M Metamodel

Finally, in [5] we have described the architecture technique, the features and the capability of ZCMGenerator to enhance the portability and reusability of existing platforms. This framework uses XML technology, which could be easily used to generate and communicate ZCM models between all the stakeholders whose are responsible for developing the ZCP platforms without considering the diversity of technological platforms. Figure 3 below describes the four steps involved in this framework.

- Analyzing: in this step, we define and describe the requirements of the users;
- Modelling: the previous requirements are transformed into a model under the following structure: schools, categories, sub-categories and wealth;
- Developing: this step consists of developing the Zakat Calculation Functions (ZCF), which allow associating the values of wealth in the XML files in order to Zakat calculation;
- Generating: this step allows generating the ZCM models in XML format in accordance with our ZC2M metamodel. It allows also generating the ZCR reports using the generated ZCM models.

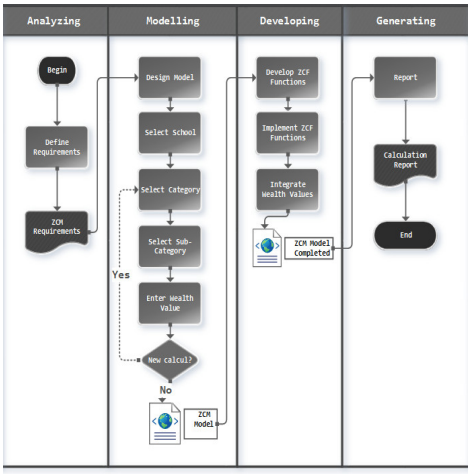


Figure 3. ZCMGenerator business processes

This paper discusses the design, implementation and performance of ZCMGenerator framework. This later is a metamodel based tool for modelling and generating of ZCM models based on the requirements expressed by users. In fact, ZCMGenerator is more than one tool of generation of ZCM models, it offers a

lot of functionality to calculate the amount owed to pay, to generate the ZCR reports and to send these reports and models by email. Another key concept of ZCMGenerator is the strong association between ZCF functions and the generated XML files. It permits through the mechanisms of transformation to define and generate the ZCM models in XML format and the ZCR reports in XHTM format. It integrates the XML technology in order to ensure the portability, reusability and interoperability.

The rest of this paper is organized as follows. Section II describes detailed analysis of various existing ZCP platforms. Section III describes in details functional and technical characteristics, business processes and development process of the ZCMGenerator. Section IV presents discussion and comparative study. Finally, section V discusses our conclusions and future work.

2 Related Work

During the past few years there has been a remarkable evolution in the world of ZCP platforms. These later are a great source for understanding and calculating Zakat. However, most of them are still having their weaknesses and limitations. Nowadays, the design and the development of a ZCP platform impose new requirements as a result of the diversity of technologies with specifics characteristic. In this context, several approaches have been proposed and used to design and develop such platform.

As technology evolves, several approaches have been proposed and multiple ZCP have been developed. In [6], a platform that calculates the Zakat for mobile phone users with a GPRS connection has been developed using J2ME technology. The IZAMD platform is based on client-server approach. The server side is just various web pages to get the prices of gold, silver and stocks. The mobile device side allows setting the Zakat configuration. This platform covers all school of jurisprudence. However, it is limited to the specific categories of wealth.

Al-Riyami et al. [7] have applied expert system technology in the field of Zakat to assist Muslims in the decision making of identifying the rules of making Zakat and to assist in complex calculations. The ZES platform was developed using a freeware rule-based shell called eXpertise2Go. In this platform, the user has the freedom of choice, either one type or as much as he desires to calculate in an understandable and clear manner. However, it does not state the source of the information.

Mohammad et al. [8] have also developed an expert system for the domain of Zakat. The ESs platform consists of two main components, the knowledge base uses facts to represent knowledge and the inference engine executes rules upon these facts to provide responses. This platform is very helpful in calculating Zakat. However, it cannot answer to questions of all Muslims belonging to different Islamic schools.

Harun et al. [9] have proposed an ontology-based approach for developing the Zakat management system. The proposed ontology has been implemented in OWL and modelled with the Protégé tool. It was designed for interoperability of systems and will make the process of developing the Zakat management system faster. However, it only covers two main processes that are collection and distribution.

Abdul Hamid & Kasirun [10] have been developed Personal Islamic Asset Management System (PERISA) using object-oriented approach by implementing Rational Unified Process (RUP) model. This platform is user friendly and easy to use for various backgrounds of users. However, its features can still be improved by integrating online payment, registration and online financial report.

Fenty et al. [11] have applied mobile application development life cycle approach in the development of Z2MWA platform using JQuery Mobile Framework for mobile users. This platform can be accessed by any platform and delivers speed, stability and an excellent cross-browser experience for web mobile visitor. However, it is limited to one school of jurisprudence and does not cover all categories of wealth.

Noorul et al. [12] have developed Muslim Android Application for Zakat Selangor (MAAZS) using agile development approach, RUP model, PHP and JQuery language for Android mobile platform. This platform is useful for Muslims users as it combine the Zakat information, Zakat calculation, Skim Berkat which is Zakat monthly deduction, Fidyah and Kifayah calculation in one application. However, it is developed especially for Selangor residents.

Ahmad et al. [13] have proposed the requirements analysis approach using the activity theory to develop a Muslim Android Application (M2A) basing on Android platform. This platform is equipped with Zakat calculator, information about Zakat, Skim Berkat and a special feature to find the nearest Zakat counter. However, it is limited to one school of jurisprudence.

Imam & Usman [14] have developed SICZ platform using Waterfall model. This later is done in a systematic and sequential approach starting from system level requirements and then headed to the stage of the analysis, design, coding, and implementation. This platform was developed to help the community in general and in particular the users of Android-based Smartphone in term of calculating Zakat. However, it does not cover all categories of wealth.

Atunnisa et al. [15] have developed ABACZ platform using prototype method. This later is an attractive idea for complicated and large systems for which there is no manual process or existing system to help determining the requirements. However, it is usually not complete and many of the details are not built in it. This platform can present the calculation of Zakat Al-Mal with easy to use and practical. However, it does not cover all types of wealth and all schools of jurisprudence.

We have now covered the main characteristics, limitations and advantages of the existing ZCP platforms. All these platforms are very useful and helpful to perform property calculations. However, our analysis shows that although multiple ZCP using different approaches have already been proposed which help in calculating the amount owed to pay. These ZCP do not provide complete requirements of Zakat calculation. For example, the most of these platforms are limited to one school of jurisprudence and are poorly implemented with limited calculation categories. Moreover, the vast majority of them do not meet the new requirements that have emerged in the field of software engineering such as portability, reusability and interoperability. Also, this analysis shows that there are two main categories for such platforms viz; desktop-based and web-based platforms and nine main groups of ZCP development approaches. Like any new platform, the adoption of these platforms by Muslims is not free from issues. To fulfil the needs of current ZCP issues and to overcome the aforementioned drawbacks a new modernized ZCP must be designed to fulfil the Zakat calculation requirements.

3 Framework and Implementation

3.1 Functional and technical characteristics

The framework ZCMGenerator is the first and innovative web framework dedicated to the Zakat domain for the modeling and generation of ZCM models, it is developed using the most advanced languages such

as JEE and XML in web-based environment. It is designed to be used in several technological platforms and to allow the reuse, portability and scalability of ZCM models. This framework makes the reuse of these models possible by describing with a metamodel the elements and their relations constituting a ZCM model. Thus, the portability and extensibility of these models are made possible by the ability to instantiate them in a format based on XML technology. As the first web framework dedicated to the Zakat domain, ZCMGenerator makes it very simple to generate simple ZCM models, while providing a high-level framework for more complex ZCM models. The technical features and benefits of this framework are many and varied viz;

A user-friendly interface that makes it easy for users to design their ZCM models, this interface is designed using Tiles technology, which is a framework designed to easily allow the creation of web application pages with a consistent look and feel;

Compliance with standards, models generated in XML format conform to W3C (World Wide Web Consortium) standards, these files are verified and tested by W3C accredited validators;

A point of communication with other platforms, the technology of ZCMGenerator is based on the use of XML files and especially in the generation of ZCM models, to guarantee a channel of communication and cooperation with other platforms;

It is a web application developed using the JEE and the XML technology in order to be available to a maximum number of users and to be consultable with any Web-browser. It can be accessed from computer, Tablet, Smartphone or any other device; and so on.

3.2 ZCMGenerator business processes

A business process is the combination of a set of activities with a structure describing their logical order. Its objective is to produce a desired result [16]. It can be visualized as a flowchart or as a process matrix of a sequence of activities. The business process of ZCMGenerator framework involves three important processes.

3.2.1

Process of modelling and generating of ZCM Models: The modelling of the ZCM models is carried out by respecting our ZC2M metamodel. The new models generated are strictly in accordance with the SCW (School-Category-Wealth) structure. This process makes it possible to transform the needs expressed by the users into a model. Each defined specification is associated with an element of the SCW structure. In this process, the users start modelling by choosing an Islamic school of their belonging, according to their choice, ZCMGenerator will help them to construct the remaining elements of their model (categories, Sub-categories and wealth). At the technical level, due to the importance of the XML technology in the field of software engineering, ZCMGenerator instantiates the new models in this format.

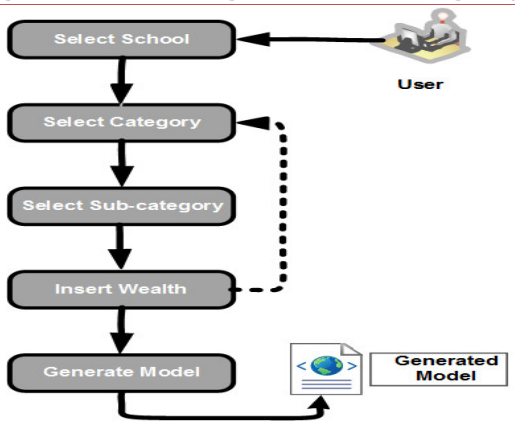


Figure 4. Process of modelling and generating of ZCM models

3.2.2

Process of calculation and generating ZCR reports: Zakat calculation is done through the generated models. To this end, the ZCF functions affect the values to Zakatable wealth after their implementation in the environment of the development. Thus, ZCMGenerator permits a numerical representation of the characteristics composing these models. Indeed, the reports generation process allows extracting directly from the generated models all the desired information and to export them in open formats.

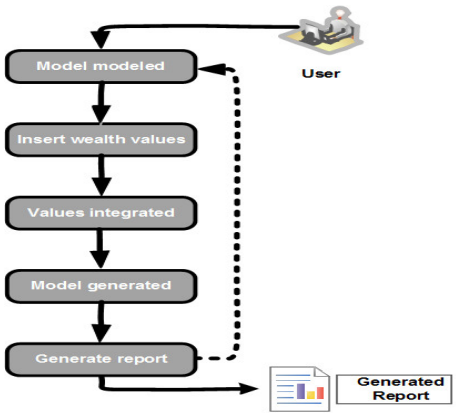


Figure 5. Process of calculation and generating ZCR reports

The generated reports are developed based on the XSLT (eXtensible Stylesheet Language Transformations) technology. Indeed, W3C proposed the XSLT standard to develop transformations of XML documents. The approach defined in this standard allows specifying transformations of XML documents using Template. This later consists of an applicability clause and a set of instructions. The operating principle of XSLT is to browse a given XML document at the input of a transformation in order to detect which templates can be applied to it. When a Template is detected as being applicable, the set of instructions that it defines is executed.

3.2.3

Process of sending and receiving e-mail notifications: The process of sending notifications allows users to send the reports and models generated by e-mail with the objective of allowing the reuse of these models

in future projects. To this end, we used the e-mail management standard of JEE technology, JavaMail API. This later allows sending and receiving e-mail and handling messages. Its purpose is to be easy to use and to provide flexibility that allows it to evolve and to remain as independent as possible of the protocols used. To send or receive messages, JavaMail uses different protocols such as SMTP, IMAP or POP. Simple Mail Transfer Protocol (SMTP) is the standard messaging protocol. The Post Office Protocol (POP) allows retrieving mail from a remote server. Internet Message Access Protocol (IMAP) is an alternative to POP offering more possibilities such as concurrent access management, multiple mailboxes and more. In our case, we have used the SMTP protocol due to its simplicity and efficiency.

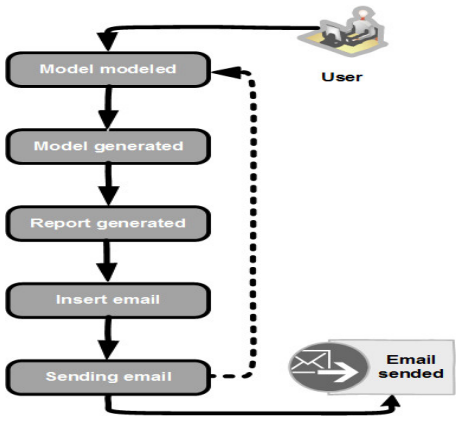


Figure 6. Process of sending email notifications

3.3 ZCMGenerator development process

The development process of our ZCMGenerator framework consists of four main phases viz; domain analysis, architectural design, implementation and validation phase. We describe these phases in detail below.

3.3.1

Domain analysis: In this first phase, we have developed a solid knowledge of Zakat domain. To this end, we first tried to identify the commonalities points of view and differences among Islamic schools exist on this subject. Then, we studied the different ZCP platforms to get an overview of what is proposed. Through a comparative analysis, we have been able to describe in detail the important characteristics, advantages and limitations of them. Subsequently, after analyzing these platforms, we have been able to define and design a specific model (PSM) corresponding to each platform. After analyzing the structures of these PSM models and from the comparison between their structures, we have proposed a general model (PIM). This later is obtained by integrating the common features of all PSM models in a common model. The general model obtained describes all the business classes, their attributes and operations and their relationships in a UML class diagram. It allows us to generate all existing ZCP models according to the requirements of the designer. In order to propose a new architecture and new features for ZCP platforms, we have enriched this general model. The result of this work is a new improved PIM model that we call ZC2M metamodel. This later is a high level abstraction metamodel which we used to create a database in the XML files format allowing the representation of the knowledge of this domain. The ZC2M metamodel is a sufficiently generic metamodel to model the structure of the ZCM models and their relationships.

3.3.2

Architectural design: The architecture of software describes its major components, their relationships and how they interact with each other. It serves as a blueprint for software and provides an abstraction to manage the software complexity and to establish a communication and coordination mechanism among components. The primary goal of the architecture is to identify requirements that affect the structure of the software [17]. In such architecture, we have implemented the MVC 2 (Model-View-Controller) architecture using the Struts 2 framework, which is a standard for developing well-architected web applications based on the MVC design paradigm [18]. In MVC2 architecture there is only one controller that receives all requests for the application and is responsible for taking appropriate action in response to each request. Then, we have identified the components and their collaboration that we need to implement our framework. Subsequently, we have determined the internal and external components necessary to ensure the proper functioning of this framework, among these important components we mention the following main APIs.

JDOM (Java Document Object Model), API for reading, writing, and manipulating XML from within Java code;

JavaMail, API for sending and receiving email via SMTP, POP3 and IMAP protocols;

XSLT: API for transforming XML documents into other documents;

JUnit: API for writing and running unit tests; and so on.

The ZCMGenerator is a set of software components structured and cooperated with each other to define the basic functions of this framework. Our development approach ensures reuse, portability, scalability, extensibility and more.

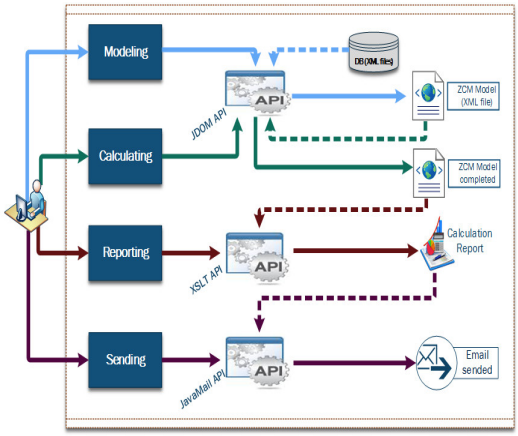


Figure 7. Technical architecture of ZCMGenerator

3.3.3

Implementation: In this phase, we have developed each element of the software architecture and implemented the basic functionalities of our framework. We present in the following sub-sections the graphical interfaces representing these functionalities.

Creation of a new ZCM model:. The ZCMGenerator allows users, using its user-friendly interfaces, to create their own ZCM models, through these interfaces, users can choose their school of belonging,

categories, subcategories and wealth, according to the SCW (School-Category-Wealth) structure which is inevitably required by the respect of our ZC2M metamodel. Figure 8 below allows users to create a specific model corresponding to the Maliki School and to Livestock Zakat. To do this, the user should choose the Maliki School, the category "Livestock", the sub-categories "Camels, Cattle and Goat-Sheep" and, thereafter, the wealth corresponding to each selected sub-category appears.

Figure 8 (1) corresponds to the interface of the Islamic schools;

Figure 8 (2) corresponds to the interface of categories of wealth;

Figure 8 (3) corresponds to the interface of the sub-categories;

Figure 8 (4) corresponds to the interface of the wealth;

Figure 8 (5) allows to user to add other wealth.

After the user has defined the ZCM model corresponding to the Maliki School and to Livestock Zakat, he has also the possibility to extend this model by adding as much wealth as he wants to calculate in a clear and comprehensible way. For instance, the user can extend the generated model by adding Zakat on Income. For this end, he should therefore choose the category Income, the sub-category (Salaries, professionals, etc.) and in the stage of the wealth corresponding to each sub-category, he can enter the corresponding values in each property field if he wants to calculate the Zakat after the generation of his model. By clicking on the "Download your Zakat Calculation Model" button (Fig. 9), the ZCMGenerator allows the automatic generation of this model in XML format (Fig. 10).

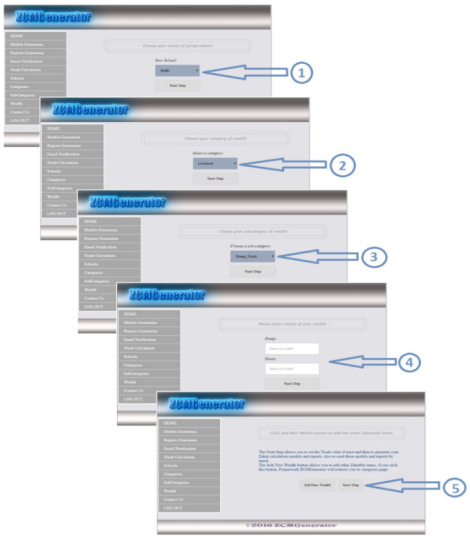


Figure 8. ZCM model modeling

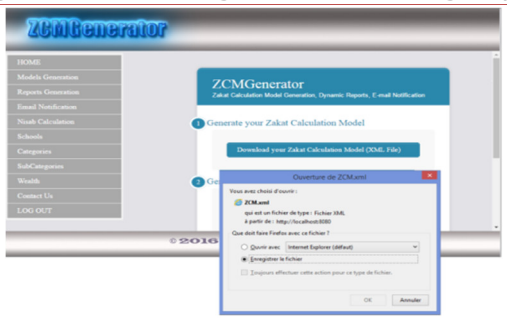


Figure 9 Instantiation of ZCM model in XML format

```xml
<?xml version="1.0" encoding="UTF-8"?>
<School name="Maliki">
  <Category name="Livestock">
    <SubCategory name="Sheep_Goats" Nissab ="40" Rate="1">
      <Wealth name="Sheep">
        <value> </value>
      </Wealth>
      <Wealth name="Goats">
        <value> </value>
      </Wealth>
    </SubCategory>
    <SubCategory name="Camels" Nissab ="5" Rate="1">
      <Wealth name="Camels">
        <value> </value>
      </Wealth>
    </SubCategory>
    <SubCategory name="Cows_buffaloes" Nissab ="30" Rate="1">
      <Wealth name="Cows">
        <value> </value>
      </Wealth>
      <Wealth name="buffaloes">
        <value> </value>
      </Wealth>
    </SubCategory>
  </Category>
</School>
```

Figure 10. The generated ZCM Model

Calculation and generation of ZCR reports: In the previous sub-section, we have shown that our framework allows users, using its user-friendly interfaces, to create their own ZCM models. In this second sub-section, we see how to calculate the Zakat from these models. To do this, in the wealth step, the users have to enter the values corresponding to the fields in the form above (Figure 8 (4)) and as explained in the previous sub-section, they have the possibility to add as much wealth as they want to calculate, for example, in addition to the Zakat on Livestock, they can add Zakat on Income. To this end, in the wealth step corresponding to the Income category, users have to enter the values corresponding to each field. After the user has entered the data requested by our framework, by clicking the "Download your Zakat Calculation Report" button (Fig. 11), the ZCMGenerator allows the automatic generation of ZCR report in XHTML format (Fig. 12).

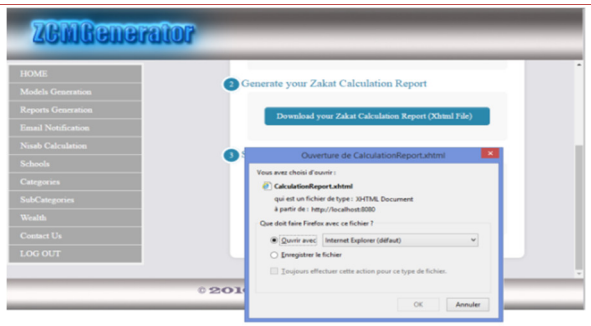
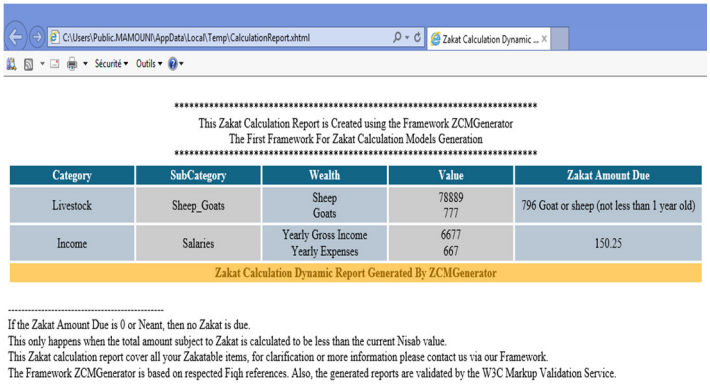


Figure 11 Generating ZCR report



This Zakat Calculation Report is Created using the Framework ZCMGenerator
The First Framework For Zakat Calculation Models Generation

Category	SubCategory	Wealth	Value	Zakat Amount Due
Livestock	Sheep_Goats	Sheep Goats	78889 777	796 Goat or sheep (not less than 1 year old)
Income	Salaries	Yearly Gross Income Yearly Expenses	6677 667	150.25

Zakat Calculation Dynamic Report Generated By ZCMGenerator

If the Zakat Amount Due is 0 or Neant, then no Zakat is due.
This only happens when the total amount subject to Zakat is calculated to be less than the current Nisab value.
This Zakat calculation report cover all your Zakatable items, for clarification or more information please contact us via our Framework.
The Framework ZCMGenerator is based on respected Fiqh references. Also, the generated reports are validated by the W3C Markup Validation Service.

Figure 12. The generated ZCR report

Sending ZCR reports by email: In the two previous sub-sections, we have seen how to create ZCM models, how to calculate the Zakat from these models and how to generate the ZCR reports. In this sub-section we will see how to send these reports by email. To do this, simply enter a recipient email, for example, “mamouni.abdelaziz@gmail.com” and then click on the "Send E-mail" button (Fig. 13). After clicking this later button, the user receives an email containing his ZCR report. Below, we present the result of sending this report by email (Fig. 14).

Figure 13. Sending ZCR reports by email

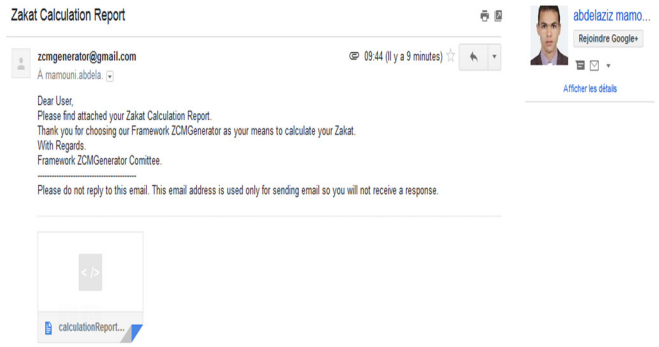


Figure 14 Message received by the user

3.4 Validation of results

In this step we have verified all the business functions for which ZCMGenerator has been designed. To do this, we have created a database in the form of XML files based on our ZC2M metamodel and on the actual data. Subsequently, we have created an XSD schema corresponding to these XML documents using a tool proposed by XMLGrid.net. Then, for the validation of the results of the generation of ZCM models in XML format, as well as the generation of the ZCR reports in XHTM format, we have used external tools that positively tested the conformity of these generated files to the standards of W3C, we present below the tools that we have used to validate our results.

The XmlGrid.net Generator service provides an XSD schema generation service from an XML document. Our XML documents check result was well formed by this service (Fig. 15) and the generated XML schema is presented as an XML file format (Fig. 16).

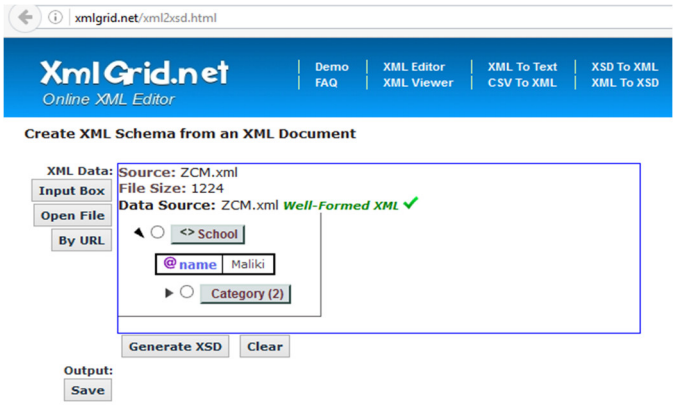


Figure 15. Result of XMLGrid Generator

Figure 16. The generated XSD schema

The Freeformatter Validator provides a validation service to validate an XML file against a specified XSD schema. The result of the generation of ZCM model in XML format was validated by this validator (Fig. 17).

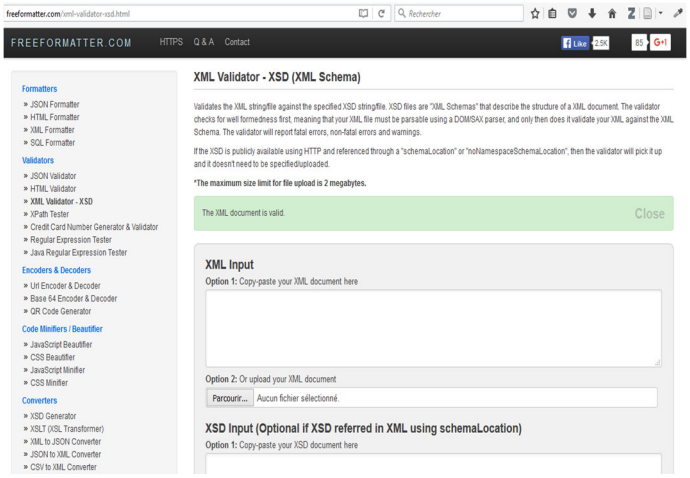


Figure 17. Result of Freeformatter Validator

The W3C validator presents an XHTM document validation service. The result of the generation of ZCR report in XHTML format was well validated by this validator (Fig. 18).

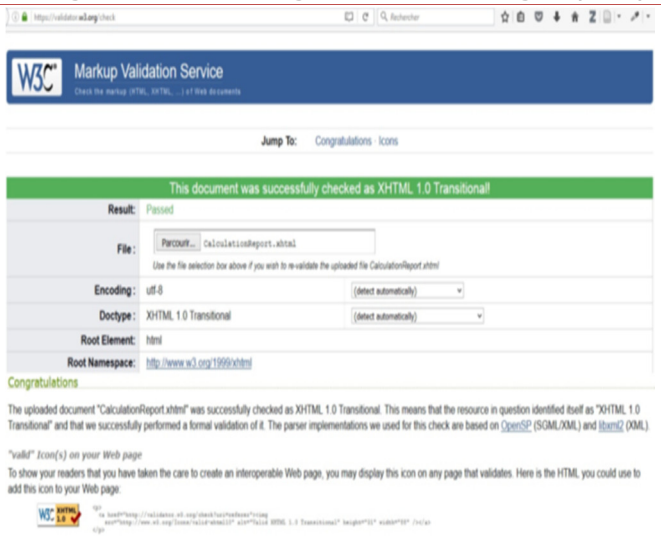


Figure 18. Result of W3C XHTML Validator

4 Discussion and Comparative Study

In the second section, we have analyzed the different existing ZCP platforms and covered the main characteristics, limitations and advantages of these platforms. This analysis shows that although multiple ZCP using different approaches have already been proposed, they do not provide complete requirements of Zakat calculation, for instance, the most of them are limited to one school of jurisprudence and are poorly implemented with limited calculation categories. Moreover, the vast majority of them do not meet the new requirements that have emerged in the field of software engineering. Furthermore, they do not meet the needs of almost all Muslims living in modern societies. In this forth section we compare the characteristics of our framework with those of the existing ZCP and we discuss the reasons for which the framework proposed presents the best results. This comparison is presented in the form of a table (Table I) and a graph (Fig. 19), the table lists the characteristics of each ZCP while the graph shows the frequency of the characteristics in each ZCP. These frequencies are ranked in descending order.

Table 1 Comparative study: ZCMGenerator Vs Existing ZCP Platforms

Criteria	ZCP Platforms											
	ZZMWA	*IZAMD*	*ZES*	*ZCE*	*ZMS*	*ESs*	*PERISA*	*SICZ*	*MAAZS*	*ABACZ*	*M2A*	*ZCMGenerator*
ZCM Modeling												X
ZCM Generating												X
ZCR Generating												X

ZCR Sending	X											X
Cover all wealth												X
Cover all schools		X										X
Full & partial modes	X	X	X		X							X
Nissab calculation	X								X	X		X
Ease of use	X	X	X	X	X	X	X	X	X	X	X	X
Multi-platforms	X						X					X
Zakat information	X	X	X	X	X	X	X	X	X	X	X	X
Meet the new requirements of software engineering												X

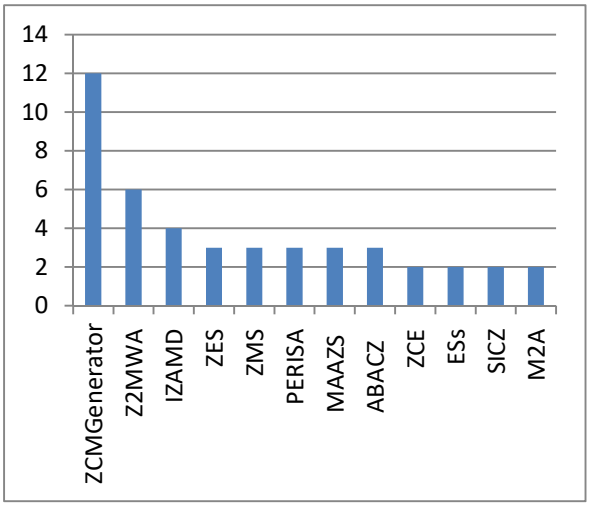


Figure 19. A comparison of the characteristics of ZCP platforms

As shown in the Table I above, there is no platform to allow modelling and generation of ZCM models while enabling the users to calculate their Zakat on any Zakatable wealth according to the principles of their jurisprudence school, whereas our framework is able to overcome these limitations by providing features to support the Muslims in calculating their Zakat according to the rules recognized by their school of belonging, whatever the desired Zakatable wealth. Thus, it allows generating models adaptable to the recent technologies and specific to each of these schools. More specifically, the Table I and Figure 19 above show the qualities of the proposed framework compared to existing platforms. From these results,

we deduce that: ZCMGenerator is the first framework allowing the modelling and the generation of ZCM models adaptable to the new technologies and specific to each school of jurisprudence. These models are extensible, reusable, portable and can be communicated to all technological platforms regardless of technical specifications; It allows the automatic generation of ZCR reports in the form of XHTML files, while the other solutions only allow to display the calculation result on their user interfaces; It not only has the ability to generate the ZCR reports in the XHTM format and the ZCM models in the XML format but also the ability to send these reports and models by email in order to allow their reusability in other projects; It is designed to work on multiple platforms, ZCMGenerator is a web application developed to be accessed by any hardware containing a web browser; It enables Zakat to be calculated on any Zakatable wealth regardless of the school of jurisprudence, whereas most of the existing platforms are designed for certain types of wealth and for a specific school; It offers two modes of calculation, the full mode calculates everything about Zakat, whereas the partial mode allows calculating the Zakat on a particular type of wealth, also, the user has the freedom of choice, either one type or as much as he desires to calculate in an understandable and clear manner; It meets the new requirements that have emerged in the field of software engineering such as portability, reusability and interoperability. Indeed, the proposed framework not only takes advantage of the MDA approach but also integrates other important aspect of portability, interoperability and reusability that is XML technology. The XML brings a number of powerful capabilities to information modelling such as heterogeneity, extensibility, flexibility, internationalization and many others. Then, by automating of the generation of ZCM models in this format, our tool addresses many issues such as reusability, portability, interoperability, flexibility and so on; The technology used during ZCMGenerator development takes into account the possibility to add new schools of jurisprudence and new Zakatable wealth, while increasing the ZCMGenerator performance. To promote high-performance and high scalability of this framework, the design pattern strategy and XML technology were used. Actually, for implementing the ZCF functions, we have used the strategy pattern. To this end, we have created a strategy interface defining an action, context class which uses a strategy and concrete strategy classes implementing the strategy interface, whereas, for implementing these functions, the existing platforms use multiple if-else statements, which allow a choice to be made between two possible alternatives. This approach is not recommended since it does not make it possible to handle more data while reducing the time required for handling more user requests, while the strategy is a better design pattern to handle such a case, which replaces if-else statements with a polymorphic call. It enables an algorithm's behaviour to be selected at runtime. The experimental results indicate that this pattern has significantly reduced the amount of time required to calculate the amount owed to pay than the calculation by using multiple if-else statements. Thus, as we have already seen above, the ZC2M metamodel was used to create a database in the XML files format allowing the representation of the knowledge of the Zakat domain. Each XML file conforms to the same XSD schema and depends on the specific school. We have created several XML files in order to allow handling the small amounts of data, when files are small it is much quicker and easier to parse, whereas, the existing platforms store all of their data in a database which needs to be read in completely to update any one field. Our framework has then the capability to handle a growing amount of data and the potential to be enlarged in order to accommodate that growth.

5 Conclusion

This paper provides the first and innovative web framework dedicated to the Zakat domain for the modelling and generation of ZCM models, ZCMGenerator was developed using the most advanced languages such as JEE and XML in web-based environment. It was designed to be used in several technological platforms and to allow the reuse, portability and scalability of ZCM models. It makes the reuse of these models possible by describing with a metamodel the elements and their relations constituting a ZCM model. Thus, the portability and extensibility of these models are made possible by the ability to instantiate them in a format based on XML technology. As the first web framework dedicated to the Zakat domain, ZCMGenerator makes it very simple to generate simple ZCM models, while providing a high-level framework for more complex ZCM models. For the validation of the results of the generation of ZCM models in XML format, as well as the generation of the ZCR reports in XHTM format, we have used external tools that positively tested the conformity of these models to the standards of W3C. Experimental results show that, compared with traditional ZCP platforms, ZCMGenerator is more scalable and high-performance.

REFERENCES

[1] A. Mamouni, A. Marzak, and Z. A. Haddad, 'ZCMGProcess: ZCM-based process for the Zakat calculation models generation (ZCMGenerator a framework supporting this approach)', in 2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA), 2016, pp. 1–6.

[2] K. A. Ali, 'XML Technology', 1983.

[3] A. Mamouni, A. Marzak, Z. Al Haddad, and Y. Boukouchi, 'Meta-Model for Zakat Calculation Platforms (ZCP) Based on the ADM Approach', Jun-2016.

[4] A. Mamouni, A. Marzak, Z. Al Haddad, and A. Belangour, 'A meta-model for automation of the deduction of judgments relating to Zakat', Dec-2015.

[5] A. Mamouni, A. Marzak, Z. Al Haddad, and Y. Boukouchi, 'ZCMGenerator: Framework for Generating Zakat Calculation Models based on the MDA Approach', Sep-2016.

[6] A. Huraimel, M. J. Zemerly, and A. Al-Hammadi, 'Islamic Zakah Application for Mobile Devices', presented at the The 3rd International Conference on Information Technology ICIT 2007, AL-Zaytoonah University, Amman, Jordan, 2007.

[7] A. Al-Riyami, A. A.-H. K. Al-Amri, and K. A. Al-Busaidi, 'Zakat Expert System', Vol. One, p. 31, 2014.

[8] A. Mohammad, A. Ahmad, H. Hassan, S. Rashid, and Z. Maamar, 'Design and Development of a Zakat Expert System', presented at the The 3rd Annual Undergraduate Research Conference on Applied Computing (URC 2011), DUBAI, UNITED ARAB EMIRATES, 2011, p. 35.

[9] H. Harun, N. Nordin, and A. Hussain, 'Ontology of Zakat Management System', in transfering, Managing and Maintaining knowledge for nation capcity development, Langkawi, Malaysia, 2008.

[10] N. F. I. Abdul Hamid and Z. M. Kasirun, 'PERISA: A PERSONAL ISLAMIC ASSET MANAGEMENT SYSTEM USING OBJECT-ORIENTED APPROACH', presented at the Proceeding of the 2nd International Conference on Informatics, Hilton Petaling Jaya Hotel, Kuala Lumpur, Malaysia, 2007, pp. 37–42.

[11] E. Fenty, K. Hulliyah, and M. Ekafitri, 'Applying mobile application development life cycle in the development of Zakat maal mobile web application using JQuery mobile framework', in Cyber and IT Service Management (CITSM), 2014 International Conference on, 2014, pp. 89–92.

[12] Noorul, Izzatthol, and Akhbarie, 'Muslim Android Application for Zakat Selangor (MAAZS)', Kuala Lumpur, Malysia, 2012.

[13] N. A. N. Ahmad, N. I. Akhbariee, and M. Hafizuddeen, 'Requirements analysis of android application using activity theory: A case study', in Information and Communication Technology (ICoICT), 2013 International Conference of, 2013, pp. 145–149.

[14] H. F. IMAM and S. Usman, 'SISTEM INFORMASI PERHITUNGAN ZAKAT BERBASIS ANDROID', Skripsi Fak. Ilmu Komput., 2015.

[15] R. Atunnisa, E. Satria, and R. Cahyana, 'PENGEMBANGAN APLIKASI ZAKAT BERBASIS ANDROID MENGGUNAKAN METODE PROTOTYPE', J. Algoritma, vol. 11, no. 1, 2015.

[16] R. S. Aguilar-Saven, 'Business process modelling: Review and framework', Int. J. Prod. Econ., vol. 90, no. 2, pp. 129–149, 2004.

[17] Tutorialspoint, 'SOFTWARE ARCHITECTURE & DESIGN INTRODUCTION'. 2013.

[18] M. Jibitesh, Software Engineering. India: Pearson Education India, 2011.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

A Model Driven Approach for Modeling and Generating PHP CodeIgniter based Applications

Karim Arrhioui, Samir Mbarki, Oualid Betari, Sarra Roubi, Mohammed Erramdani
MISC Laboratory, Faculty of Sciences, IbnTofail University, Kenitra, Morocco
MATSI Laboratory, Superior School of Technology, Mohammed First University, Oujda, Morocco
arr.karim@gmail.com, mbarkisamir@hotmail.com, beta.oualid@gmail.com, roubi.sarra@gmail.com, m.erramdani@gmail.com

ABSTRACT

During the last decade, web development industry has grown exponentially. Models have been introduced as a solution to face the challenge of both business and technology changes. In this article, we present a Model Driven based approach concerning the design of CodeIgniter based web applications. We describe a meta model of this framework and we also specify a set of transformations to generate the application's source code taking into account the MVC (Model-View-Controller) architecture of CodeIgniter. In this approach, the PHP framework meta model is considered as a platform Specific model (PSM). Its instances are used as inputs to generate the source code through transformation rules carried out by Acceleo. This proposal is validated through the use of our approach to generate CRUD (Create, Read, Update and Delete) applications.

Keywords-CodeIgniter; Model Driven Architecture; Model-View-Controller; PHP; Platform Specific Model

1 Introduction

The express development of web based application has affected the coding methodology. This points toward a higher need of sustainability and maintainability. Besides, PHP is a web scripting language which is used in dynamic interactive web development. It is a general-purpose and an open source tool that requires minimal setup [1]. Furthermore, PHP has become one of the most powerful programing languages for developing web applications. To deal with the problems caused by the increasing projects' complexity, several techniques for programming in PHP such as Procedural PHP coding and Object Oriented Programming have been proposed. Resulting this rapid development, several frameworks such as CodeIgniter PHP framework have emerged to facilitate the development tasks. Indeed, as mentioned in [2],they have been improved and "become handy tools for developers to build complex applications efficiently".

In order to benefit from using these frameworks and also handle frequent changes, the Object Management Group (OMG) proposed the Model Driven Architecture as a solution. This approach supports changing business rules in different application domains by providing an open approach to manage the challenge of interoperability, by using OMG's established modeling standards that are: Unified Modeling Language (UML) and Meta-Object Facility (MOF) [3] [4].

DOI: 10.14738/tmlai.54.3189
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3189

In this paper, we consider the union of the solutions offered by the introduction of the PHP frameworks and the use of MDE (Model Driven Engineering) in developing applications, as we will apply a Model Driven approach to model the CodeIgniter PHP framework and generate CRUD applications based on this framework.

The present paper is structured as follows: In section 2 we outline the MDE principles. Section 3 presents the concepts of CodeIgniter and the MVC (Model-View-Controller) pattern. Section 4 introduces our MDA approach for generating CRUD applications. A case study is presented in Section 5. Finally, section 6 concludes the work and offers further perspectives.

2 Model Driven Architecture

The OMG has introduced the Model Driven Architecture that is based on Modeling and transformation to generate the code and presented a formal statement with several tools and approaches [3]. This architecture focuses on creating models with a high level of abstraction, and promotes transforming the models according to given defined rules [5].

Today's systems are constantly changing and highly networked. In order to face these challenges, MDA provides a platform independence architecture that assures portability and cross-platform Interoperability [3].

MDA consists of three general types of models, structured into three basic layers: Computation Independent Model (CIM), Platform Independent Model (PIM), and Platform Specific Model (PSM). The three models are defined as follow:

- CIM: This model describes the system's functionalities with a high level of abstraction. It is seen as a business model, as it uses a vocabulary that is familiar to the subject matter experts. It presents what the system is supposed to do, and hides structure details and implementation [6].
- PIM: This model defines the concept of the system without showing the specific details of a target platform. It exhibits a sufficient degree of independence so as to enable its mapping to one or more platforms.
- PSM: It describes the deleted details and characteristics of PIM. It also provides platform specific details that should be considered to implement the system.

The reason for the above model organization is to develop models of the systems' business logic independently from the platforms of execution, then to transform these models automatically to models dependent of the platforms. The complexity of the platforms does no longer appear in the business logic models but it is found in the transformation [7].

The Model Driven development using UML approach requires several steps: at first building the CIM that acquires user requirements. Then, according to this CIM, a PIM is built. Next, the proposed PIM is mapped into one or more PSMs. This type of transition from CIM to PIM and PIM to PSM is called Model To Model (M2M) transformation. Finally, the code of the target platform is generated from the PSM instance. This transition is called Model To Text (M2T) transformation [5] [7] [8]. Fig. 1 shows how the transformations are done [9]:

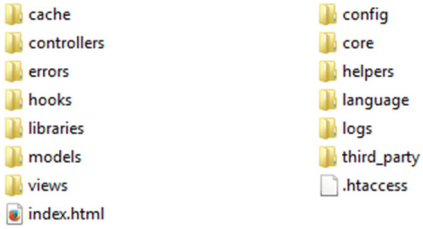


Figure 1. Model Driven Architecture Layers.

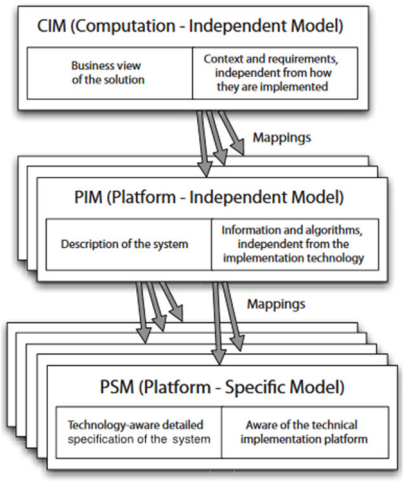


Figure 2. Structure of CodeIgniter applications

CodeIgniter is a PHP Application Development Framework based on a well-structured architecture. It aims to provide necessary tools such as helpers and libraries to implement common tasks. Thus, project development becomes much easier and faster, and developers don't have to write all the code from scratch [2] [10].

CodeIgniter is based on the MVC development approach, MVC is a design pattern that structures the software by separating application logic from presentation [11]. Indeed, PHP scripting related to business elements is separated from web pages [12].

- The Model contains the business logic of the application. It gathers functions related to data accessing and third-party services.
- The View is composed of user interface elements such as HTML, CSS and JavaScript files. In CodeIgniter, a view can be a web page, a fragment of a page, a RSS page...
- The Controller connects views, models and any needed resources to process and respond to the user request. It is the point of entry that instantiates the required views and models [13].

3 Model Driven Approach for generating CRUD applications

The development of web applications has been improved by the integration of frameworks, CodeIgniter being one of these frameworks allows the development of PHP web applications.Using CodeIgniter for the development of web applications, it is necessary to respect a precise structure as shown in Fig. 2:

That's why when we initiated the development of a meta-model for this framework, it was necessary to study this structure and raise the level of abstraction.

The proposed meta model in this paper is the PSM that describes the CodeIgniter PHP framework. This meta-model is used to define instances of models that describe web applications. These instances will be the inputs of the transformation engine that will be developed using Acceleo that respects an approach by template. The primary benefit of our approach is that the long task of coding a CRUD web application is done in a systematic, well-structured and standard way by using MDA principles.

3.1 The proposed CodeIgniter meta model

The developed meta model of CodeIgniter framework, shown in Fig. 3, is represented by the model, view and controller packages. Each package contains specific meta classes according to the MVC pattern.

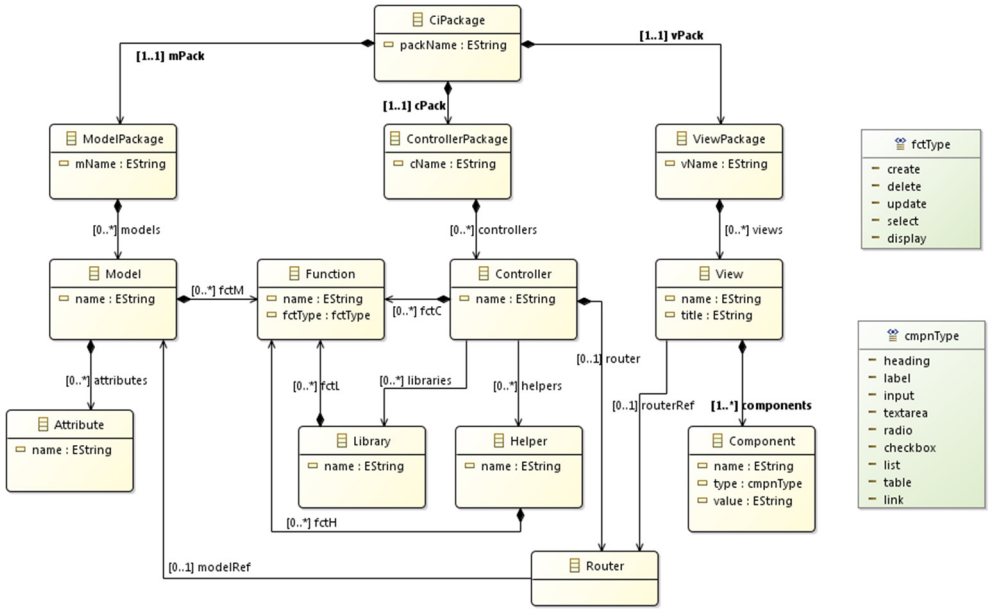


Figure 3. CodeIgniter meta model.

- CiPackage: expresses the concept of package that includes the entire elements of the model; the model, the controller and the view, that are themselves regrouped, respectively, in ModelPackage, ControllerPackage and ViewPackage.
- A helper or a library is a set of functions that can be called from the controller, the view or the model. It is a PHP file that contains functions grouped by theme [12].
- View: contains the components of the View that will be defined bellow.
- Controller: the intermediary between the model and the view. It contains all the functions to manage each user action [14].
- Component: it is a graphical element such as input, text area and heading.

The View/Controller layers are responsible for describing the structure and content of views, while the navigation flow is ensured through the controller's specific functions that are connected to the specified services from the model layer. This layer gathers all the business parts of the application.

3.2 The Model to Text transformation rules

Once the application has been sufficiently modelled, the code generation procedure follows the Model To Text transformation (M2T) to get, in our case, a CodeIgniter application's source code using Acceleo [8]. This transformation follows the template approach; thus, we have developed the needed templates for code generation taking as an input a CodeIgniter meta model instance. The main module for code generation is shownbellow:

```
[comment encoding = UTF-8 /]
[modulegenerateCI('http://cimd/1.0')]
[templatepublicgenerateElement(aCiPackage : CiPackage)]
[comment@main/]
        [for(model : Model|aCiPackage.mPack.model)]
                [generateModels(model)/]
        [/for]
        [for(controller : Controller|aCiPackage.cPack.controller)]
        [generateControllers(controller)/]
        [/for]
        [for(view : View|aCiPackage.vPack.views)]
                        [generateViews(view)/]
        [/for]
[/template]
[templatepublicgenerateControllers(controller : Controller)]
        [file ('/controllers/'+controller.name+'.php',false,'UTF-8')]
<?php
class[controller.name/] extends CI_Controller {
                public function index()
{   $data['['/]'title'[']'/] = "Add [controller.name/]";
            $data['['/]'heading'[']'/] = "-- Add [

controller.name/] --";
            $this->load->view('[controller.name/]_view', $data);
   }
                [for(fct : Function|controller.fctC)]
                        [generateFct(fct)/]
                [/for]
}
?>
        [/file]
[/template]
```

4 Case Study

CRUD operations are often the most commonly used in web applications. In addition, CodeIgniter offers the possibility to manage these operations through its predefined structure. So the approach we propose comes to simplify and further automate this process which converges towards these objectives and allows to generate these CRUD operations automatically in terms of source code.

```php
<?php
class Customer extends CI_Controller {

    public function index()
    {
        $data['title'] = "Add Customer";
        $data['heading'] = "-- Add Customer --";

        $this->load->view('Customer_view', $data);
    }

    public function addCustomer()
    {
        $this->load->helper('form');
        $this->load->model('CustomerModel');

        $Customer_array = array(
        'id'        => NULL,
        'name'      => $this->input->post('name')
      , 'email'     => $this->input->post('email')
      , 'adress'        => $this->input->post('adress')
        );
```

Figure 4. The input model of Customer CRUD operations

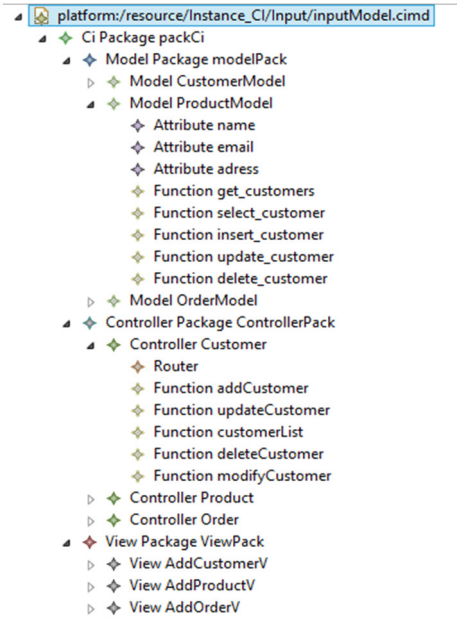


Figure 5

Once the application is modeled,using Acceleo, the code generation procedure follows the templates developed for the M2T engine. Fig. 5 below is an excerpt from the generated code:

To add a new customer, the user should provide a name, an address and an email then submit the form to the customer controller. Fig. 6 shows the generated views containing forms for adding both customer and product. Once the form is submitted, the specific controller calls a function in order to add the new element

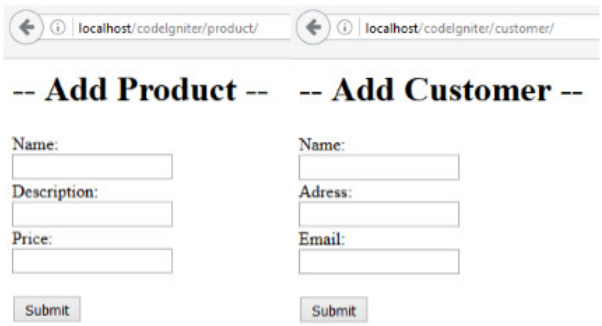
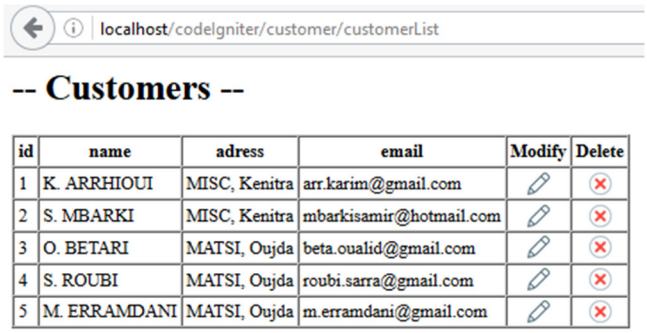


Figure 7. Generated forms for adding customer and product

To list all customers, the customer controller loads the customer model, calls “get_cutomers” function and then sends the result to the customerList view as show in Fig. 7:



id	name	adress	email	Modify	Delete
1	K. ARRHIOUI	MISC, Kenitra	arr.karim@gmail.com		
2	S. MBARKI	MISC, Kenitra	mbarkisamir@hotmail.com		
3	O. BETARI	MATSI, Oujda	beta.oualid@gmail.com		
4	S. ROUBI	MATSI, Oujda	roubi.sarra@gmail.com		
5	M. ERRAMDANI	MATSI, Oujda	m.erramdani@gmail.com		

Figure 8. Customers list from the database

5 Conclusion and Perspectives

In this paper, we proposed an approach that allows modeling web applications based on the CodeIgniter PHP framework. In order to achieve this end, we applied the model driven approach principles to generate these applications. This approach using modeling and transformation process provides advantages in improving applications portability and quality, while minimizing cost and time. This approach has demonstrated its efficiency through enabling the building of a complete meta model that describes MVC CodeIgniter application, then, generating its source code.

The form validation, using jQuery Validation Plugin and CodeIgniter form validation library, will be the subject of a study thereafter. Afterwards, we will consider applying this solution to other PHP frameworks in order to propose a comparative study.

REFERENCES

[1] S. Bergmann, and G. Kniesel, "GAP: generic aspects for PHP," in Proc. EWAS'06, 2006.

[2] O.Betari, M. Erramdani, S. Roubi, K. Arrhioui, and S. Mbarki, "Model transformations in the MOF meta-modeling architecture: from UML to codeIgniter PHP framework," Europe and MENA Cooperation Advances in Information and Communication Technologies, vol. 520, pp. 227-234, 2016.

[3] Object Management Group (OMG), MDA Guide 2.0. http://www.omg.org/cgi-bin/doc?ormsc/14-06-01

[4] Executive Overview, Model Driven Architecturehttp://www.omg.org/mda/executive_overview.htm

[5] S. Kherraf, É. Lefebvre, and W. Suryn, "Transformation from CIM to PIM using patterns and archetypes," 19th Australian Conference on Software Engineering (aswec 2008), Perth, WA, pp. 338-346, 2008.

[6] S. Roubi, M. Erramdani, and S. Mbarki, "Modeling and generating graphical user interface for MVC rich internet application using a model driven approach," 2016 International Conference on Information Technology for Organizations Development (IT4OD), Fez, pp. 1-6, 2016.

[7] X. Blanc, MDA en Action: Ingénierie Logicielle Guidée par les Modèles, Eyrolles, 2005.

[8] S. Roubi, M. Erramdani, S. Mbarki, "Generating graphical user interfaces based on model driven engineering,"in International Review on Computers and Software (IRECOS), vol. 10, pp. 520-528, 2015.

[9] M. Brambilla, J. Cabot and M. Wimmer, Model-Driven Software Engineering in Practice, Morgan & Claypool, 2012.

[10] E. Orr, and Y. Zadik, Programming with CodeIgniter MVC, Ed. Birmingham, UK: Packt Publishing, 2013.

[11] R. Foster, CodeIgniter 2 Cookbook, Ed. Birmingham, UK: Packt Publishing, 2013.

[12] CodeIgniter Documentation website. https://www.codeigniter.com/docs/

[13] C. Pitt, Pro PHP MVC, Apress, 2012.

[14] M. Brambilla and A. Origgi, "MVC-Webflow: an AJAX tool for online modeling of MVC-2 web applications," 2008 Eighth International Conference on Web Engineering, pp. 344-349, 2008.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Building An Automatic Speech Recognition System for Home Automation

[1]Mohamed Aboulkhir, [1]Samira Khoulji, [2]Reda Jourani, ML Kerkeb
[1]Spatial Remote Sensing -Signal Processing Applied Maths &Computer Science Decision Support Laboratory, Abdelmalek Essaadi University, National School of Applied Sciences, 93000,Tetouan,Morocco.
[2]Polydisciplinary Faculty of Tetouan, Abdelmalek Essaadi University, Tetouan, Morocco.
amohamed.aboulkhir@gmail.com, bkhouljisamira@gmail.com,creda.jourani@yahoo.fr, kerkebml@gmail.com

ABSTRACT

This paper presents a study on automatic speech recognition (ASR) systems applied to home automation. So a detailed study of the architecture of speech recognition systems was carried out. The objective is to select a speech recognition software that must operate in remote speech conditions and in a noisy environment. The proposed system is using an ASR toolkit called Kaldi, which must communicate as an open platform communication (OPC) client developed in C++, with any home automation system. The latter behaves like an OPC server.

Keywords : Speech recognition, acoustic model, language model, HMM, n-gram, domotics,Kaldi.

1 Introduction

Speech is the most natural mode of communication. Through it, we can give voice to our thoughts. We can use it to express opinions, ideas, feelings, desires or to exchange, transmit, request information. And today we do not just use it to communicate with other humans, but also with machines.

Speech recognition is the technique that allows the analysis of sounds picked up by a microphone to transcribe them into a series of words that can be used by machines. Since its appearance in the 1950, automatic speech recognition has been constantly improved. Today the applications of speech recognition are very diverse and each system has its own architecture and mode of operation. The wider the field of application, the greater the recognition models must be (in order to understand spontaneous discourses and the diversity of the speakers)[1].

In our study, we will be interested in the automatic recognition of speech applied to home automation. Indeed, it is the goal pursued by the smart home which is a residence equipped with computer technology to assist its inhabitants in the various situations of the domestic life as well in terms of comfort as that of security. Automatic Speech Recognition could be an essential contribution to the detection of abnormal situations, which is an essential part of a home surveillance system [2].

The paper is organized as follows: we start by describing the components of a speech recognition system. This is followed by describing components of many speech recognition software and testing their

DOI: 10.14738/tmlai.54.3190
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3190

performance in order to choose the one, which will subsequently be integrated into a home automation system.

2 Automatic Speech Recognition

2.1 Overview

A speech recognition system is intended to associate a sequence of words with a sequence of acoustic observations. Thus, from the sequence of acoustic observations X, this system searches for the sequence of words Ŵ which maximizes the probability P (W | X)

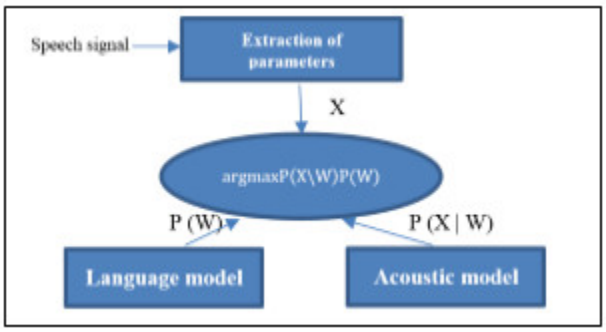


Figure. 1 Schematics of the operation of a speech recognition system

which is the probability of emission of W knowing X[3]. The sequence of words Ŵ must then maximize equation:

$$\hat{W} = argmaxP\left(\frac{W}{X}\right) \quad (1)$$

Applying the Bayes rule, we obtain the formula

$$\hat{W} = argmax\frac{P(X\backslash W)P(W)}{P(X)} \quad (2)$$

Since $P(X)$ is constant, then:

$$\hat{W} = argmaxP(X\backslash W)P(W) \quad (3)$$

Two types of probabilistic models are used to search for the most probable sequence of words: an acoustic model that provides the value of P (X | W), and a language model that provides the value of P (W). Fig. 1 shows a general schematic diagram of the operation of an automatic speech recognition system[4].

2.2 Feature Extraction

As can be seen in Fig. 1, the speech signal cannot be directly transformed into hypotheses of word sequences. The extraction of its parameters is an important step since it must determine the relevant characteristics of the signal. This extraction can be done using multiple techniques, the most well-known ones being parametric analysis, using the LPC (Linear Predictive Coding) method, the cepstral analysis, with for example the MFCC (Mel-scale Frequency Cepstral Coefficients), or the PLP (Perceptual Linear Prediction) technique. These different methods make it possible to extract characteristic coefficients for each frame. This extraction then makes it possible to obtain the sequence of acoustic observations X [5].

2.3 Acoustic Model

The acoustic model is a statistical model that estimates the probability that a phoneme has generated a certain sequence of acoustic parameters. A wide variety of acoustic parameter sequences are observed for each phoneme due to all variations related to speaker diversity, age, gender, dialect, state of health, emotional state. The most widely used methods for acoustic modeling are models based on hidden Markov models (HMM) or deep neuron networks (DNN)[6].

2.4 Language Model

Language models are processes that estimate the probabilities of the different word sequences P (W). These models are used to memorize sequences of words from a textual corpus of learning. In the context of speech recognition, language models serve to guide and constrain research among alternative word hypotheses[7]. The most commonly used language models are n-gram models

2.5 Evaluation of speech recognition systems

In order to evaluate several speech recognition systems, they should be compared on the same test data. Conventionally, these systems are evaluated in terms of word-error rates[8]. The WER takes into account the errors of:

- Substitution: recognized word in place of a word of manual transcription.
- Insertion: Recognized word inserted in relation to the reference transcription.
- Deletion: word of forgotten reference in the hypothesis provided by the speech recognition system.

The WER is expressed by the formula:

$$WER = \frac{substitutions+insertions+suppressions}{\text{number of words in the reference}} \quad (4)$$

3 Automatic speech recognition tool

There are several open-source software for automatic speech recognition (ASR). Notable among these are HTK, Julius (both written in C), Sphinx-4 (written in Java), RWTH ASR toolkit and Kaldi (both written in C++)[9].

In the first phase, we selected an ASR software with the characteristics adapted to the construction of our system, or the least satisfactory to the best of all these needs. After we will describe the features of automatic speech recognition supported by this software.

3.1 ASR Software: Kaldi

To advance our study we choose the voice recognition software called Kaldi. It is an open-source toolkit for speech recognition written in C++ and licensed under the Apache License v2.0.[10]. This choice is motivated by:

- Kaldi have modern and flexible code written in C++ that is easy to understand, modify and extend.
- Open license: The code is licensed under Apache v2.0, which is one of the least restrictive licenses available.

- Extensible design: The algorithms are developed in the most generic form possible. This will allow us to easily integrate our home automation.
- Extensive linear algebra Support: it include a matrixlibrary that wraps standard routines.
- Complete recipes: it make available complete recipes for building speech recognition systems that work from widely available databases.
- Performance: Kaldi outperforms all the other recognition toolkits[11].

3.2 Overview of Kaldi

Kaldi is a speech recognition toolkit consisting of a library, command lineprograms and scripts for acoustic modelling [12].The architecture of Kaldi, as described in Figure 2, consists of the following modules:

- External Libraries: Kaldi depends on two external libraries that are also freely available. One is OpenFst for the finite-state framework (FST), and the other is numerical algebra libraries such as BLAS "Basic Linear Algebra Subroutines" and LAPACK "Linear Algebra PACKage"[13].
- Kaldi C++ Library: It contains all the functionalities and different modes of a speech recognition system developed by C ++, which are then called from a scripting language for building and running a speech recognizer.
- Kaldi C++ Executables Scripts

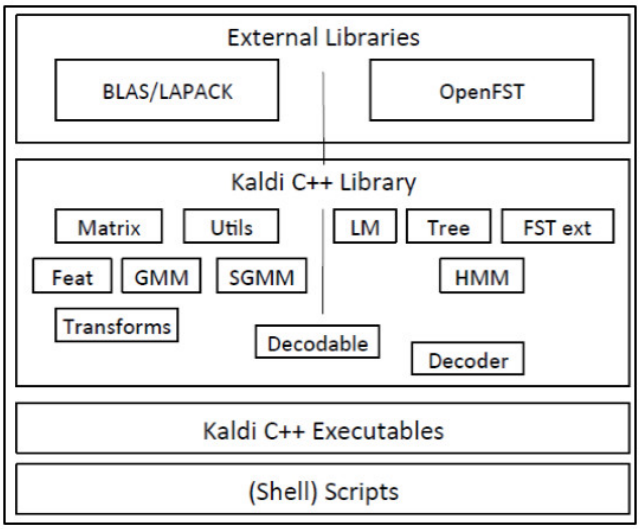


Figure. 2 – Kaldi architecture

As described in the second part of this document, a speech recognition system consists of three modules: Parameter extraction, acoustic model, language model. The Kaldi library integrates these three elements and proposes the following codes:

- For the extraction of parameters: the Kaldi code aims at creating standard functions of MFCC and PLP, setting reasonable default values, but leaving people the opportunity to change the values.
- For the acoustic model: Kaldi supports conventional acoustic models such as GMM, SGMM, HMM and DNN, it is also extensible to new types of models.

```
HRESULT hr;

hr = CLSIDFromProgID(lpwSeverName, &clsid);

hr = CoCreateInstance(clsid, NULL, CLSCTX_ALL, IID_IUnknown, (LPVOID *)&pUnkn);

// Get connection-pointer

hr = pUnkn->QueryInterface(IID_IOPCServer, (LPVOID*)&m_pOpcServer);
```

- For the language models: Kaldi allows to use any language model that can be represented as a FST. Therefore,it supports the most used n-gram model[14].

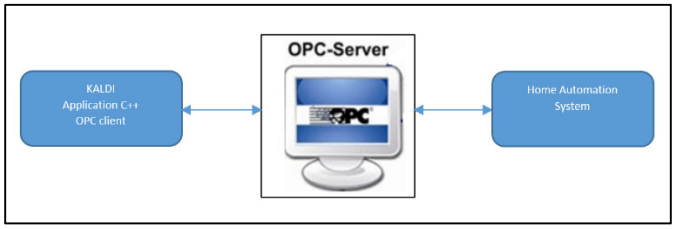
4 Kaldi for Domotics

The aim of our work is to integrate our chosen speech recognition software Kaldi, to a home automation system. For this, we will first propose an architecture of integration, which will be the object of several tests in order to qualify the performances of our model.

4.1 Integration Architecture

In a first step, we propose a communication architecture between Kaldi and a home automation system. Our proposal was based on an architecture based on the OPC client / server technique (Figure 3)[15]. This choice is motivated by:

- The C ++ language supports OPC. Thus, Kaldi can be configured as an OPC client [16].
- The client / server communication can be integrated with the most home automation systems[17].




```
// Write by IOPCSyncIO.

hr = m_pOPCGroup->QueryInterface(IID_IOPCSyncIO, (LPVOID*)&pOPCSync);
```

Figure. 3 – Integration architecture

4.2 OPC Routines for kaldi

We will detail in this section the steps developed on Kaldi for the implementation of the client / server OPC communication between our ASR software (called client) and home automation system through its OPC server.The first step is to connect to OPC Server;the following code will allow this connection:

The second step is to Create OPC Group Object and Add Tags

The last step is to configure read and write to and from an OPC server. In our case, we will only configure the write:

```
// Add OPCGroup
HRESULT hr = m_pOpcServer->AddGroup(GroupName.AllocSysString(), Active,
                UpdateRate, (OPCHANDLE)pNewGroup, &Bias, &Deadband,
                LocaleID, &(pNewGroup->m_hServerHandle), &Rate,
                IID_IOPCGroupStateMgt, (LPUNKNOWN*)&pInterface);

// Add OPCItems
OPCITEMDEF* idef = new OPCITEMDEF[nPoints];
for(DWORD i=0; i<nPoints; i++)
{
  CItemObj* pItemObj = TempList->GetAt(pos);
  CString csName = pItemObj->m_Name;
  idef[i].szItemID = csName.AllocSysString();
  idef[i].dwBlobSize = 0;
  idef[i].pBlob = NULL;
  idef[i].bActive = TRUE;
  idef[i].hClient = (OPCHANDLE)pItemObj->m_hClientHandle;
  idef[i].szAccessPath = AccessPath.AllocSysString();
  idef[i].vtRequestedDataType = VT_EMPTY;
  TempList->GetNext(pos);
}
hr = m_pOPCGroup->QueryInterface(IID_IOPCItemMgt, (LPVOID*)&m_pOPCItem);
hr = m_pOPCItem->AddItems(nPoints, idef, &pResults, &pErrors);
```

5 Conclusion

After the presentation of the state of the art of automatic speech recognition systems,we described the design of Kaldi, a free and open-source speech recognition toolkit. It supports a wide range of methods for extracting parameters, acoustic models and language models. We were also able to configure Kaldi as an OPC client in order to be able to integrate it into a home automation system through its OPC server.Future work will concentrate on implementing this integration in order to test the robustness of our system.

REFERENCES

[1] Allauzen A. et Gauvain J.-L., Construction automatique du vocabulaire d'un système de transcription, dans Journées d'Étude sur le Parole (JEP)

[2] Michel Vacher. Analyse sonore et multimodale dans le domaine de l'assistance à domicile. Intelligence artificielle [cs.AI]. Université de Grenoble, 2011.

[3] Richard Dufour. Transcription automatique de la parole spontanée. Informatique [cs]. Université du Maine, 2010.

[4] Mohamed Bouallegue. L'analyse factorielle pour la modélisation acoustique des systèmes dereconnaissance de la parole. Autre [cs.OH]. Université d'Avignon, 2013.

[5] Insect sound recognition based on mfcc and pnn. In Multimedia and Signal Processing (CMSP), 2011 International Conference IEEE

[6] Fethi Bougares. Attelage de systèmes de transcription automatique de la parole. Ordinateur et societé [cs.CY]. Université du Maine, 2012.

[7] Panagiota Karanasou. Phonemic variability and confusability in pronunciation modeling forautomatic speech recognition. Other [cs.OH]. Université Paris Sud - Paris XI, 2013.

[8] Ngoc-Tien Le, Christophe Servan, Benjamin Lecouteux, Laurent Besacier. Better Evaluatio of ASR in Speech Translation Context Using Word Embeddings. Interspeech 2016.

[9] AMAN F., VACHER M., PORTET F., DUCLOT W. & LECOUTEUX B. (2016). CirdoX : an On/Off-line Multisource Speech and Sound Analysis Software. In LREC 2016.

[10] Madikeri, S., Dey, S., Motlicek, P., & Ferras, M. (2016). Implementation of the standard i-vector system for the kaldi speech recognition toolkit (No. EPFL-REPORT-223041). Idiap.

[11] Gaida, C., Lange, P., Petrick, R., Proba, P., Malatawy, A., & Suendermann-Oeft, D. (2014). Comparing open-source speech recognition toolkits. Tech. Rep., DHBW Stuttgart.

[12] The Kaldi Speech Recognition Toolkit, Povey Daniel, Ghoshal Arnab, Boulianne, GillesBurget, LukasGlembek, OndrejGoel, Nagendra, Hannemann, MirkoMotlicek Petr, Qian Yanmin, Schwarz Petr, Silovsky Jan, Stemmer Georg and Vesely Karel, Idiap-RR-04-2012

[13] Povey, D., Hannemann, M., Boulianne, G., Burget, L., Ghoshal, A., Janda, M., ... & Riedhammer, K. (2012, March). Generating exact lattices in the WFST framework. In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference. IEEE.

[14] C. Allauzen, M. Riley, J. Schalkwyk, W. Skut, and M. Mohri, “OpenFst:a general and efficient weighted finite-state transducer library,” in Proc.CIAA, 2007.

[15] Olivier Passalacqua, Eric Benoit, Marc-Philippe Huget, Patrice Moreaux. INTEGRATINGOPC DATA INTO GSN INFRASTRUCTURES. IADIS International Conference APPLIEDCOMPUTING 2008

[16] Zheng, L., & Nakagawa, H. (2002, August). OPC (OLE for process control) specification and its developments. In SICE 2002. Proceedings of the 41st SICE Annual Conference (Vol. 2, pp. 917-920). IEEE.

[17] Topalis, E., Orphanos, G., Koubias, S., & Papadopoulos, G. (2000). A generic network management architecture targeted to support home automation networks and home internet connectivity. IEEE Transactions on Consumer Electronics, 46(1), 44-51.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Smart Classroom Environment Via IoT in Basic and Secondary Education

Hicham EL Mrabet, Abdelaziz Ait Moussa
Department of Computer Sciences, Faculty of Sciences, Mohammed First University
Oujda, Morocco
hi.elmrabet@gmail.com; a_aitmoussa@yahoo.fr

ABSTRACT

The rapid evolution of information and communication technologies (ICT) has created a new paradigm of the Internet, known as the Internet of Things (IoT), which represents a group of connected objects Using wireless links (WiFi, RF, NFC, Zigbee…)to exchange information between objects and persons. The Internet of Things has changed the way people and objects interact with each other, and education has not been immune to this change, which has created new forms of interaction between teachers and learners that helps improve the teaching and learning process.In the field of education, we can imagine that students use new technologies to carry out projects and educational activities in the classroom. In this research,the focus is on the application of Internet of Things in the Smart Classrooms.This paper presents and discusses the need of integrating the Internet of Things to improve learning at the basic and secondary school, mainly through information and communication technologies (ICT) at its disposal.

Keywords- Internet of Things, Smart Classroom, ICT, Flipped Classroom

1 Introduction

Today, several information and communication technologies (ICTs) are present throughout the world, and their use is progressing considerably in terms of education. ICTs have great power to improve the outcomes of teaching and learning. The realization of this power depends very much on how the teacher uses technology during the learning sessions, knowing thatthe education is a powerful path for social change, social mobility, it helps develop and build a new generation of our nation. In this context, smart classes play an important role in the transformation of traditional education into modern education, which offers opportunities for improving the quality of education and academic achievement, access to education and equity, through mobile devices, classes that integrates many different types of computer hardware, including computers, tablets, interactive whiteboards, smartphones, and many computer technologies that are Used for teaching purposes. When talking about intelligent classes, we must evoke the notion of the Internet of objects (IoT), which represents the universe of connected objects in all domains, such as the field of education.

The level of connectivity offered by the Internet of Things makes it possible to enrich the learning process for students all over the world, and the application of IoT in education aims to build an ecosystem where Students and teachers can gain a thorough and empirical understanding of their environments and can trigger change of learning using IoT

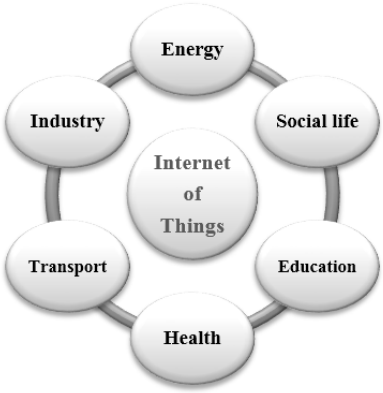
DOI: 10.14738/tmlai.54.3191
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3191

2 Literature review

Technology is among the elements that strongly influences everyday life, and currently, the Internet of Things has the power to change our world, a paradigm shift in technology that will invade many areas of our daily life.

2.1 Internet of things

The Internet of Things represents a network that enables digital identities and physical objects to be identified directly via standardized electronic identification systems and wireless devices (WSN, NFC, Zigbee, RF…), and thus to be able to retrieve, store, transfer and process discontinuity between physical and virtual worlds. The Internet of Things has the power to change our world, a paradigm shift in technology that will invade multiple domains, and the following figure outlines the most useful areas for IoT.



2.2 Internet of things in education

The education system offers permanent opportunities for intelligent learning, content, pedagogy or values to convey. Generally, it is asserted that pedagogy is the way to transfer knowledge between the teacher and the learners. At all times, the teacher must update his knowledge transfer methods and techniques.

The integration of IoT gives the learner the possibility of being also the mediator of knowledge in order to reinvest in the class the knowledge and the strategies acquired outside the class and that the teacher is definitely the more the only mediator of knowledge, he will feel valued and will be more motivated.

3 Proposed Work

In this section, we present related work, which focuses on the smart classroom environment. Through this study, it should show how we can improve the learning process based on related educational approaches.we propose new learning approach, knowing that the evolution of the research is mainly motivated by the taking into account of the new technologies of IoT in parallel to the teaching pedagogical approaches.

3.1 The structure of smart learning process via IoT

Integration of IoT as a new actor in intelligent learning can facilitate the interaction between people (learners and teachers) and objects (Devices, RFID Tag ...) in the school environment. This interaction

means that objects can communicate with each other and with people who are in these educational environments.

Smart learning process lets learner complete their coursework on their mobile devices (Tablets, Smartphones,Laptop, connected objects…) and learners have complete access to their real-time class and collaborative learning from their mobile device.

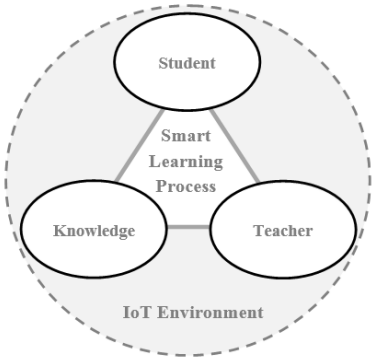
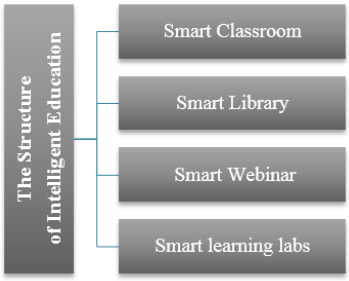


Figure 1 IoT in Smart learning

The structure of smartlearning and a management system include the following aspects:



3.2 The IoT-Learning approach

Our IoT-Learning approach based on connected objects in basic and secondary school, is represented by a pedagogical triangle, modified and surrounded by a sphere representing the universe of IoT.The IoT-Learning model can be represented by the following diagram

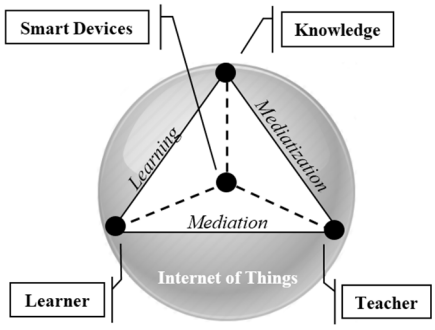


Figure 2. The IoT-Learning approach

In our model the 'Devices' pole of the classic triangleis replaced by 'Smart Devices'.This is due to the logic of the fact that the pedagogical situation using ICT is essentially based on access to information. The

teacher plays a dual role. It governs, on the one hand, the process of mediation, by which it ensures the pedagogical relationship in order to facilitate the acquisition of knowledge by the learner. It then fulfills a pedagogical function of management, interactive regulation of learning events. On the other hand, it assumes the process of medialization, by which it is responsible for transforming a part of collective knowledge into useful information for learning.In this sense, books, digital textbooks, posters and interactive boards, as well as the training system, the virtual campus, etc., appear as the result of the medialization process.

By adopting the power of ICT and the Internet of things in educational practice, learners benefit in many ways. Technology helps them to approach all learning styles, participate interactively and sharing, making learning personal, dynamic, collaborative and up to date. In addition, learning goes beyond class. Interacting with so many everyday objects connected to the Web, students have access to unlimited course from anywhere, and anytime. The smart class system identifies each student already integrated into each student ID card. The system is expected to use IoT node transceivers, making it perfect for indoor location.

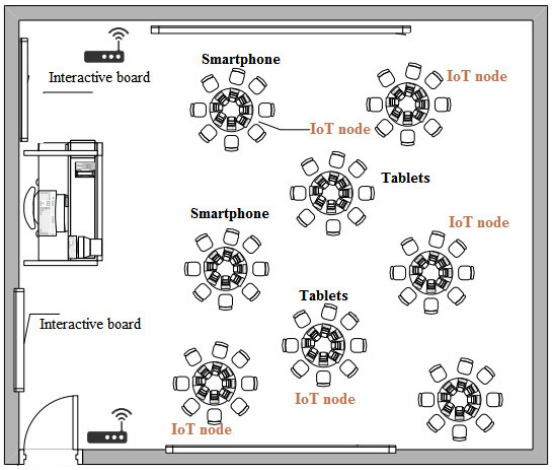


Figure 2. Architecture of the future classroom

In this case study,oursmart classroom for basic and secondary school, students will implement in all classes of the school and aims to solve the full participation rate of entire students at any period.The IoT technology integrated into the classroom is a special classroom with a precise purpose and mission can provide students with alternative methods and choices.

Research in the teaching / IOT discipline should address these three factors:

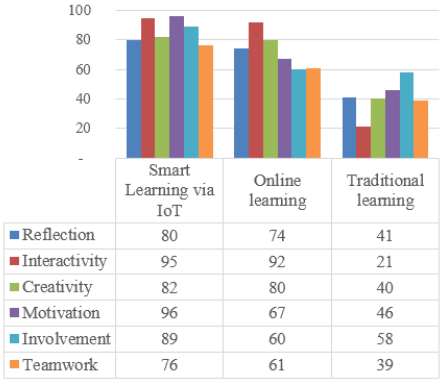
- Confidentiality
- The availability of data
- Pedagogical approaches to teaching

4 Discussion and Future Works

We presented a new framework for recommending new pedagogical approaches for integrating IoT into smart classesand the concept of collaborative learning via IoT. Besides, the proposed approach IoT-Learning cover other important dimensions, such as the impact of the IoT environment on the learners' level of knowledge and cognitive load.Future work aims to develop an intelligent environment based on an IoT authentication system that allows learners to identify themselves in intelligent classes

This global analysis blends the opinions of those who have experienced the use of connected objects in class with those who have never done so. Also, there is the risk of confusing proven benefits with hopes of profits.

The overall responses of the 100 teachers are illustrated in the following graph.



	Smart Learning via IoT	Online learning	Traditional learning
Reflection	80	74	41
Interactivity	95	92	21
Creativity	82	80	40
Motivation	96	67	46
Involvement	89	60	58
Teamwork	76	61	39

Figure 3. Comparison of different types of learning

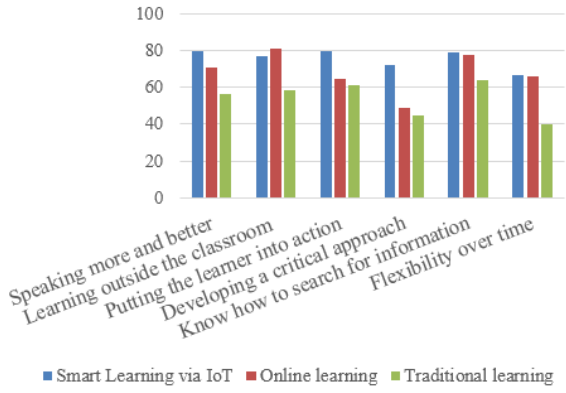


Figure 4. Advantge of Smart learning via IoT

To conclude this section, it should be noted that the overall score for observing the benefits of connected objects is significantly influenced by seniority in the career. The youngest in the teaching profession have a much more positive opinion than those who have older.

5 Conclusions

Nowadays, IoT is a smart network that connects all things to the Internet for exchanging information and communicating through exchange information and communicate through intelligent objects. There are enormous values in the adaptation of IoT throughout the education system. IoT enables the education system to be more relevant and effective. In the future, most educational institutions will gradually use IoT technologies to improve the learning process, and distance education. The new digital culture will create virtual schools that rely on Internet objects to encourage off-campus learning and expand participation in modern learning.

REFERENCES

[1] B. Jonathan, and A. Sams, “Flip your classroom: Reach every student in every class every day,” in International Society for Technology in Education, 2012.

[2] Jeannette Chinand Vic Callaghan, “Educational Living Labs: A Novel Internet-ofThings Based Approach to Teaching andResearch”, 9th IEEE International Conference on Intelligent Environments, Athens, Greece, 16 - 19 July 2013.

[3] S. Kim, S.-M. Song, and Y.-I. Yoon, “Smart learning services based on smart cloud computing,” Sensors, vol. 11, no. 8, pp. 7835–7850, 2011.

[4] T. Kim, J. Y. Cho, and B. G. Lee, “Evolution to smart learning in public education: a case study of Korean public education,” Open and Social Technologies for Networked Learning, vol. 395, pp. 170–178, 2013.

[5] Ma Guojun, Li Yueguang, “Application of IOT in Information Teaching of Ethnic Colleges”,International Conference on Information, Business and Education Technology (ICIBIT 2013).

[6] J. Valk, A.T. Rashid and L. Elder, "Using mobile phones to improve educational outcomes: An analysis of evidence from Asia",The International Review of Research in Open and Distance Learning, vol. 11, pp. 117-140, 2010.

[7] M. Saylor, “Active Learning and Virtual Worlds,” The mobile wave: how mobile intelligence will change everything, Vanguard Press. pp. 176-180, Jan. 2013.

[8] J.S. Sung, "Design of Smart Learning in Mobile Environment." IJSEIA, vol. 9, no. 12, pp. 373-380, Dec. 2015.

[9] Z.-T. Zhu, M.-H. Yu, P. Riezebos, “A research framework of smart education”. Smart Learn. Environ. 3(1), 1–17 (2016)

[10] Kim, S. (2008). Étude des représentations du personnel enseignant à l’égard de ses pratiques d’ordre technologique et pédagogique actuelles et de celles qui pourraient favoriser la mise en œuvre d’un dispositif de formation à distance à l’Institut de Technologie du Cambo. Université de Sherbrooke.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

A Comparative Study Between Operating Systems (Os) for the Internet of Things (IoT)

Aberbach Hicham, Adil Jeghal, Abdelouahed Sabrim, Hamid Tairi
LIIAN, Department of Mathematic & Computer Sciences, Sciences School, Sidi Mohammed Ben Abdellah University,
aberbachhicham@gmail.com, adil.jeghal@usmba.ac.ma, abdelouahed.sabri@gmail.com, htairi@yahoo.fr

ABSTRACT

Abstract : We describe The Internet of Things (IoT) as a network of physical objects or "things" embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data in real time with the outside world. It therefore assumes an operating system (OS) which is considered as an unavoidable point for good communication between all devices “objects”. For this purpose, this paper presents a comparative study between the popular known operating systems for internet of things . In a first step we will define in detail the advantages and disadvantages of each one , then another part of Interpretation is developed, in order to analyze the specific requirements that an OS should satisfy to be used and determine the most appropriate .This work will solve the problem of choice of operating system suitable for the Internet of things in order to incorporate it within our research team.

Keywords: Internet of things , network, physical object ,sensors,operating system.

1 Introduction

The Internet of Things (IoT) is the vision of interconnecting objects, users and entities “objects”.

Much, if not most, of the billions of intelligent devices on the Internet will be embedded systems equipped with an Operating Systems (OS) which is a system programs that manage computer resources whether tangible resources (like memory, storage, network, input/output etc.) or intangible resources (like running other computer programs as processes, providing logical ports for different network connections etc.), So it is the most important program that runs on a computer[1].

Every general-purpose computer must have an operating system to run other programs and applications.

Computer operating systems perform basic tasks, such as recognizing input from the keyboard, sending output to the display screen, keeping track of files and directories on the disk, and controlling peripheral devices such as printers. For large systems, the operating system has even greater responsibilities and powers. It is like a traffic cop — it makes sure that different programs and users running at the same time do not interfere with each other. The operating system is also responsible for security, ensuring that unauthorized users do not access the system, the most of these operating systems are used by the internet of things in several applications which are translate into many concrete uses - new or improved- significantly impacting the daily lives of individuals, businesses and communities for example[2] : Cities :

DOI: 10.14738/tmlai.54.3192
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3192

the Internet of things will allow better management of the various networks that feed our cities (water, electricity, gas, etc.) by allowing continuous real-time and precise control, Energy : power grid management will be improved by telemetry, enabling real-time management of the energy distribution infrastructure ,Transport : in this field the Internet of Things will support the current efforts around intelligent vehicles for road safety and driving assistance ,Health : in the field of health, the Internet of things will allow the deployment of personal networks for the control and monitoring of clinical signs, especially for the elderly, The industry : in the Internet the Industry of Objects will allow a total tracking of the products, from production chain, up the supply chain and distribution by overseeing the supply conditions. Because the problem of choice between different operating systems presents an obstacle in our context. The main objective of this work is to solve it in a definitive way. This work contains a detailed study of the existing operating systems for the Internet of things, and this through an evaluation of each one with its advantages and disadvantages, and the final part will be devoted to defining and examine the appropriate Operating system to be used in the Internet of Things.

2 Operating Systems

The set of operating systems designed for the internet of things are presented under the architecture shown in Figure below (Figure1) : A hard-ware layer and drivers, a second layer that contains Kernel / scheduler and network stack, The choice of Kernel is to have a complete control over everything that occurs in the system and the choice of the scheduling strategy is tightly bound to real-time and different task priorities support .Finally the last layer is for applications or user interface for supporting degree of user interaction[3] :

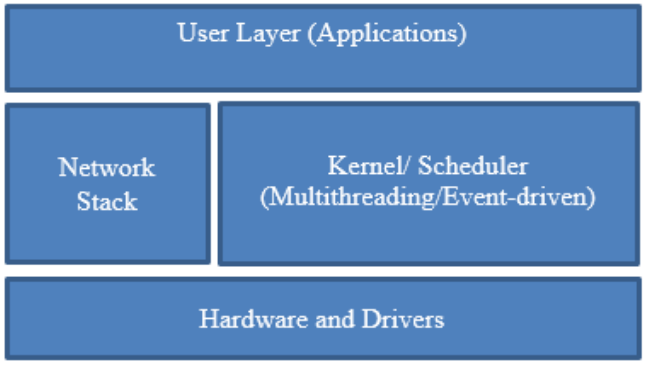


Figure1: architecture of operating systems

2.1 TinyOS

TinyOS [4] is an open source, BSD-licensed operating system designed for low-power wireless devices, such as those used in sensor networks, ubiquitous computing, personal area networks, smart buildings, and smart meters.

2.1.1 Architecture:

TinyOS follow a monolithic architecture based on a combination of components, reducing the the size of the code needed to set it up. This is in line with the constraints of Memories that are observed by sensor networks; however, the TinyOS component library is particularly complete since it includes network protocols, sensor drivers and acquisition tools of data. All of these components can be used as they are, they can also be adapted to a precise application.

2.1.2 Programming Model:

The previous versions of TinyOS do not provide multithreading support; building applications were based on the event-driven programming model. TinyOS version 2.1 supports multithreading. and these TinyOS threads are called TOS Threads[5].

2.1.3 Suported plateforms:

TinyOSTinyOS supports the following sensing platforms: Mica [6], Mica2 [6], Micaz [6] and a few others.

2.1.4 Advantages

TinyOS is a system mainly developed and Supported by the American University of Berkeley, which offers it in download Under the BSD license and monitors it. Thus, all the sources are Available for many physical targets[7].

Based on event-driven operation, TinyOS offers to the user a precise management of the sensor's consumption and makes it possible to adapt better to the Random nature of wireless communication between physical interfaces.

Compatibility with at least 9 hardware platforms, and the simplicity of adding or modifying platforms. TinyOS follows a hierarchical abstraction of the material in three layers.

The preemptive nature of an operating system specifies if this one allows the interruption of a task in progress. TinyOS does not manage this Preemption between tasks but gives priority to hardware interruptions.

Energy management: TinyOS is designed to optimally manage energy consumption: it puts the node into standby when there are no tasks to perform, and thus allows the deactivation of the radio device or its setting in ' Listening only and low energy consumption (low power listening). TinyOS also generates a small code, which decreases the energy used to store the data in the RAM.

2.1.5 Disadvantages

a) Support for Real-Time Applications

TinyOS [7]does not support real-time application operations. It provides a process planning algorithm based on priorities, but once a process is scheduled to run to completion. This may result in a missed deadline for a high priority process that enters the Preparation queue once a low priority process has been scheduled

TinyOS is not an operating system in the current sense. It does not offer, for example, a notion of multitasking, users or file system. There are no user mode and kernel mode notions. In fact, TinyOS is a set of routines made available to the programmer to simplify the development.

2.2 Contiki

Contiki [8] is an open source operating systems for sensor nodes. It was developed at the Swedish Institute of Computer Science by Dunkels et al. [3], it is a lightweight open source OS written in C for WSN sensor nodes Contiki connects tiny lowcost, low-power microcontrollers to the Internet .

Contiki is a highly portable OS and it is build around an event-driven kernel. Contiki provides preemptive multitasking that can be used at the individual process level. A typical Contiki configuration consumes 2 kilobytes of RAM and 40 kilobytes of ROM. A full Contiki installation includes features like : multitasking kernel, preemptive multithreading, proto-threads, TCP/IP networking, IPv6, a Graphical User Interface, a web browser, a personal web server, a simple telnet client, a screensaver, and virtual network computing.

2.2.1 Advantages:

a) Internet Standards : Contiki provides powerful low-power Internet communication. Contiki supports fully standard IPv6 and IPv4, along with the recent low-power wireless standards: 6lowpan, RPL, CoAP. With Contiki's ContikiMAC and sleepy routers, even wireless routers can be battery-operated.

b) Rapid Development: With Contiki, development is easy and fast: Contiki applications are written in standard C, with the Cooja simulator Contiki networks can be emulated before burned into hardware, and Instant Contiki provides an entire development environment in a single download.

c) A Selection of Hardware : Contiki runs on a range of low-power wireless devices, many of which can be easily purchased online.

d) Open Source Software: Contiki is open source software: it can be freely used both in commercial and non-commercial systems and the full source code is available.

2.2.2 Architecture

Contiki is based on a modular architecture, with respect to its kernel that follows the event driven model, but it provides threading capabilities to processes. The Contiki kernel includes a light event scheduler that dispatches events to running processes. The execution of the process is triggered by events sent by the kernel to processes or by a query mechanism.

2.2.3 Disadvantages

a) Scheduling : Contiki is an event-driven OS, therefore it does not employ any sophisticated scheduling algorithm. Events are fired to the target application as they arrive.

Contiki does not support multicast. Therefore Contiki does not provide any implementation of group management protocols such as the Internet Group Management Protocol (ICMP), or Multicast Listener Discovery (MLD) protocol

b) Possibility of Real-Time support applications : Contiki does not support deployment of real-time applications, So there is no implementation of any real-time process scheduling algorithm in this OS.

2.3 Nano-RK

Nano-RK [9] is a fully preemptive reservation-based real-time operating system (RTOS) from Carnegie Mellon University with multi-hop networking support for use in wireless sensor networks.Nano-RK currently runs on the FireFly Sensor Networking Platform as well as the MicaZ motes. It includes a light-weight embedded resource kernel (RK) with rich functionality and timing support using less than 2KB of RAM and 18KB of ROM.

Nano-RK supports fixed-priority preemptive multitasking for ensuring that task deadlines are met, along with support for CPU, network, as well as, sensor and actuator reservations. Tasks can specify their resource demands and the operating system provides timely, guaranteed and controlled access to CPU

cycles and network packets. Together these resources form virtual energy reservations that allow the OS to enforce system and task level energy budgets[8].

2.3.1 Advantages

One of the goals for Nano-RK[9] was to facilitate application developers by allowing them to work in a familiar multitasking paradigm. This results in a short learning curve, rapid application development, and improved productivity

a) Resource Sharing: For shared resources such as memory, Nano-RK provides semaphores for serialized access. To circumvent the priority inversion problem, Nano-RK provides an implementation of the Priority Ceiling algorithm. In addition, Nano-RK provides APIs to reserve system resources like CPU cycles, sensors, and network bandwidth.

b) Possibility of Real-time Applications support : Nano-RK is a real-time operating system; hence it provides rich support for real-time applications. It supports real-time processes and its offline admission control procedure guarantees to meet deadline associated with each admitted real-time process.

Nano-RK provides an implementation of real-time preemptive scheduling algorithms and tasks are scheduled using a rate monotonic scheduling algorithm. Moreover, Nano-RK provides bandwidth reservations for delay-sensitive flows and it claims to provide end-to-end delay guarantees in multi-hop wireless sensor network.

Nano-RK is a suitable OS for use in multimedia sensor networks due to its extensive support provided to real-time applications.

2.3.2 Disadvantages

Nano-RK only provides support for static memory management; it does not support dynamic memory management. In Nano-RK, both the OS and applications reside in a single address space and to the best of authors' knowledge Nano-RK does not provide any support to safeguard co-located OS and process address spaces.

Nano-RK is a preemptive multitasking OS, it needs to save the context of the current task before scheduling the new task. Saving the state of each task results in large memory consumption and frequent context switches result in reduced performance and higher energy consumption.

2.4 LiteOS

LiteOS [10] is an open source, interactive, UNIX-like operating system designed for wireless sensor networks. With the tools that come with LiteOS, you can operate one or more wireless sensor networks in a Unix-like manner, transferring data, installing programs, retrieving results, or configuring sensors. You can also develop programs for nodes, and wirelessly distribute such programs to sensor nodes.

2.4.1 Architecture

LiteOS [10] follows a modular architecture design. LiteOS is partitioned into three subsystems: LiteShell, LiteFS, and the Kernel. LiteShell is a Unix-like shell that provides support for shell commands for file management, process management, debugging, and devices.

2.4.2 Advantages

a) Scheduling

LiteOS provides an implementation of Round Robin scheduling and Priority-based scheduling. Whenever a task is added to the ready queue, the next task to be executed is chosen through priority-based scheduling. The tasks run to completion or until they request a resource that is not currently available. When a task requires a resource that is not available, the task enables interrupts and goes to sleep mode.

b) Memory Protection and Management

Inside the kernel, LiteOS supports dynamic memory allocation through the use of free functions. User applications can use these APIs to allocate and de-allocate memory at run-time. Dynamic memory grows in the opposite direction of the LiteOS stack. The dynamic memory is allocated from the unused area between the end of the kernel variables and the start of the user application memory blocks. This allows adjusting the size of dynamic memory as required by the application.

c) Communication Protocol Support

liteOS provides communication support in the form of files. LiteOS creates a file corresponding to each device on the sensor node. Similarly, it creates a file corresponding to the radio interface. Whenever there is some data that needs to be sent, the data is placed into the radio file and is afterward wirelessly transmitted. In the same manner, whenever some data arrives at the node it is placed in the radio file and is delivered to the corresponding application using the port number present in the data .

2.4.3 Disadvantages

a) Possibility of Real-time Applications support

LiteOS does not provide any implementation of networking protocols that support real-time multimedia applications. It provides a priority-based process scheduling algorithm but once a process is scheduled it runs to completion.This can result in a missed deadline of a higher priority process that enters the ready queue once a low priority process has been scheduled.

The LiteOS documentation does not provide enough detail on how system resources are shared among multiple executing threads.

2.5 FreeRTOS

“FreeRTOS [11] is a market leading RTOS from Real Time Engineers Ltd. that supports 35 architectures and received more than 113000 download during 2014. It is professionally developed, strictly quality controlled, robust, supported, and free to embed in commercial products without any requirement to expose your proprietary source code. FreeRTOS has become the de facto standard RTOS for microcontrollers by removing common objections to using free software, and in so doing, providing a truly compelling free software model.”

2.5.1 Advantages [11]

- FreeRTOS is downloaded every 260 seconds (on average).
- FreeRTOS offers lower project risks and a lower total cost of ownership than commercial alternatives because:
- It is fully supported and documented.

- Most people take products to market without ever contacting us, but with the complete peace of mind that they could opt to switch to a fully indemnified commercial license (with dedicated support) at any time.
- Some FreeRTOS ports never completely disable interrupts.
- For strict quality control purposes, and to remove all IP ownership ambiguity, official FreeRTOS code is separated from community contributions.
- FreeRTOS has a tick-less mode to directly support low power applications.
- FreeRTOS was downloaded >113000 times in 2014.
- FreeRTOS is designed to be simple and easy to use: Only 3 source files that are common to all RTOS ports, and one microcontroller specific source file are required, and its API is designed to be simple and intuitive.

2.5.2 Disadvantages

a) Limited Tasks

There are only limited tasks run at the same time and the concentration of these system are on few application to avoid errors and other task have to wait. Sometime there is no time limit of how much the waiting tasks have to wait.

b) Use heavy system resources

FreeRTOS used lot of system resources which is not as good and is also expensive.

c) Low multi-tasking

Multi-tasking is done few of times and this is the main disadvantage of FreeRTOS because this system runs few tasks and stay focused on them. So it is not best for systems which use lot of multi-threading because of poor thread priority.

d) Complex algorithms

FreeRTOS uses complex algorithms to achieve a desired output and it is very difficult to write that algorithms for a designer.

e) Device driver and interrupt signals

FreeRTOS must need specific device drivers and interrupt signals to response fast to interrupts.

f) Thread Priority

Thread priority is not good as FreeRTOS do less switching of tasks.

2.6 RIOT

RIOT [12] or the friendly Operating System for the Internet of Things is a lightweight operating system for networked systems with memory constraints, focused on devices with low power consumption for Internet of things. It is a free software, released under the GNU General Public License (LGPL)[5] , It was originally developed by the Free University of Berlin, the National Institute for Research in Computer Science and Automation (INRIA) and the University of Applied Sciences of Hamburg (HAW Hamburg). The core of RIOT is largely inherited from FireKernel1, which was originally developed for sensor networks.

2.6.1 Architecture

The microkernel architecture written in ANSI C and support for full multithreading enables a developer-friendly API. POSIX compliance is partly already available and full POSIX compliance is planned for the near future. Since RIOT is completely written in C, it also allows for the usage of C++ and the utilization of the GNU Compiler Collection (GCC) in the latest version.

2.6.2 Modularity

To ensure minimal memory usage, the system is based in a modular way. Thus, the configuration of the system can be customized to meet the particular specification. The size of the kernel itself is minimized, thus requiring only a few hundred bytes of RAM and program storage. Dependencies between the modules are reduced to an absolute minimum .

2.6.3 Advantages

RIOT [12] an operating system designed for the particular requirements of Internet of Things (IoT) scenarios. These requirements comprise a low memory footprint, high energy efficiency, real-time capabilities, a modular and configurable communication stack, and support for a wide range of lowpower devices.

RIOT provides a microkernel, utilities like cryptographic libraries, data structures (bloom filters, hash tables, priority queues), or a shell, different network stacks, and support for various microcontrollers, radio drivers, sensors, and configurations for entire platforms. There are no new programming environments.,C or C++ can be used directly with existing tools like gcc, gdb, etc, Less hardware dependent code, Supports 8-,16- and 32-bit microcontroller platforms, Energy efficiency is maintained, Less interrupt latency, so real-time capability is ensured, Multi-threading is enabled .

2.6.4 Network Stack

Supports the entire network stack of IoT (802.15.4 Zigbee, 6LoWPAN, ICMP6, Ipv6, RPL, CoAP, etc) , Both static and dynamic memory allocation, POSIX compliant (partial), All output can be seen in the terminal if hardware is not available.

RIOT is of course free and its code is available online. It is a fundamental prerequisite for developing a robust and sustainable technology and protocols on a global scale. Open source makes it possible to constantly improve a program due to a large community of contributors.

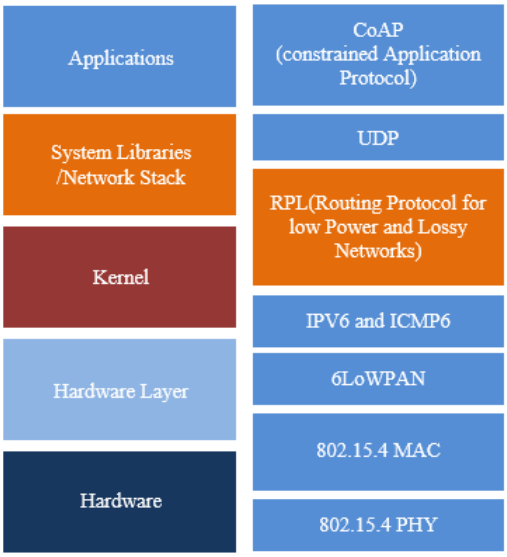


Figure 2: architecture and Driver support for RIOT

Table 1: comparison of OS's

OS	*License*	*Programming Model*	*Real-time support*	*Technologies supported*
TinyOS	***Open Source BSD***	***Event Drive,thread***	***No***	Broadcast based Routing
Contiki	***Open Source BSD***	***Protothreads and events***	***Partial***	6lowpan, RPL
NanoRK	***Open Source***	***Multithreading***	***Yes***	RTLink,U-Connect
LiteOS	***Open Source***	***Threads and Events***	***Partial***	JTAG
Free RTOS	***Open Source***	***Multithreading***	***Yes***	TCP,UDP, Ethernet, TLS
RIOT	***Open Source LGPL v2.1***	***Multithreading***	***Yes***	6LoWPAN, RPL,CoAP, UDP,TCP, CBOR, CCN-lite, OpenWSN, UBJSON

3 Comparison of IOT OS'S

As we can see in the following table (table1), the RIOT system supports the most communication technologies over other systems, supports multi-threading and real-time media with a GNU GPL license, which allows [13] the user several rights On a computer program.

1. The freedom to run the software for any purpose.
2. The freedom to study the functioning of a program and to adapt it to its needs, which requires access to source codes.
3. Freedom to redistribute copies.
4. The obligation to provide the community with modified versions.

Since Our goal is to discover and compare in a detailed way the characteristics of each operating system in order to detect the most suitable for the Internet of things. For this we have based on the characteristics

detailed in section2: OSs to conclude that the RIOT operating system is the most appropriate especially in our search axis since it is free to be run in any domain, adaptable To the needs of the user with simple access to source codes, it aims exactly to fulfill the requirements to should satisfy an operating system for internet of things .

4 Requirements of an OS for IOT

In summary the main requirements for an operating system are:

- Memory Requirements: To ensure proper memory management for an operating system the minimum memory requirement of the software must be very low. This concerns RAM as well as persistent program storage.
- Limited resources: The hardware platforms offer very limited resources so the operating system should use them eciently.
- Concurrency: The operating system should be able to handle diferent tasks at the same time.
- Flexibility: Since the requirements for deferent applications vary wildly, the operating system should be able to be flexible to handle those.
- Low Power: Energy conservation should be one of the main goals for the operating system.
- Error Free: Error free that mean it has no chances of error in performing tasks.
- Platform support: The software part of the IOT should have the ability to support the different hardware platforms, but also the ability to exploit their capabilities.
- Reliability: The operating system must function reliably because it is deployed in critical applications where the access is linked to a high costs

5 Conclusion and Future Work

In this work we have defined the notion of the Internet of things and its different fields of applications.

We have also compared the current (IoT) operating systems in a detailed way to see their advantages and disadvantages as well as their characteristics and architectures. Through this comparison we have deduced that the RIOT is the most appropriate system for the Internet of things since most of its Strengths correspond exactly to the Requirements that an operating system must satisfy to be used in the internet of things .So this article helps to better choose which operating systems to adapt for the internet of the things .This paper analyzed these parts in a single conclusion: RIOT is an operating system designed for the particular requirements of Internet of Things (IoT) Scenarios.

The future work well includes complete and compliant implementations of RIOT, on the most known platforms for the IoT.

RÉFERENCES

[1] ITU-T (2012). Overview of the Internet of things. ITU-T. Retrieved from http://www.itu.int/rec/T-REC-Y.2060-201206-I on 2016, Jan.

[2] Cisco IBSG © 2011 Cisco et/ou ses filiales. Tous droits réservés.

[3] Padmini Gaur, Mohit P. Tahiliani .Wireless Information Networking Group (WiNG)

[4] A. Dunkels, B. Gr¨onvall, T. Voigt Contiki - a Lightweight and Flexible Operating System for Tiny Networked Sensors In Proceedings of the First IEEE Workshop on Embedded Networked Sensors, Tampa, Florida, USA, 2004

[5] Klues K, Liang CJM, Paek J, Musaloiu R, Levis P, Terzis A, Govindan R. TOSThread: Thread-Safe and Non-Invasive Preemption in TinyOS. Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems; Berkeley, CA, USA. 4–6 November 2009; pp. 127–140.

[6] PtolemyProject. Availableonline http://ptolemy.eecs.berkeley.edu/ (accessed on 17 April 2011)

[7] Akyildiz IF, Melodia T, Chowdhury KR. A Survey on Wireless Multimedia Sensor Networks. Comput. Netw. 2007;51:921–960

[8] Dunkels A, Gronvall B, Voigt T. Contiki a Lightweight and Flexible Operating System for Tiny Networked Sensors. Proceedings of the 9th Annual IEEE International Conference on Local Computer Networks; Washington, DC, USA. October 2004; pp. 455–462.

[9] Eswaran A, Rowe A, Rajkumar R. Nano-RK: An Energy-Aware Resource-Centric RTOS for Sensor Networks. Proceedings of the 26th IEEE Real-Time Systems Symposium; Miami, FL, USA. 5–8 December 2005.

[10] Cao Q, Abdelzaher T, Stankovic J, He T. The LiteOS Operating System: Towards Unix Like Abstraction for Wireless Sensor Networks. Proceedings of the 7th International Conference on Information Processing in Sensor Networks (IPSN 2008); St Louis, MO, USA. 22–24 April 2008.

[11] Julien Le Sech, thirsday 5 january 2012. FreeRTOS sur ATmega328 .

[12] RIOT (2015). RIOT - The friendly Operating System for the Internet of Things. RIOT-OS.org. Retrieved from www.riot-os.org/ on 2016, Jan.

[13] Free Software Foundation https://www.gnu.org/licenses/lgpl-3.0.en.html .

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

SAAS Cloud security : Attacks and Proposed Solutions

Sail Soufiane, Bouden Halima
Laboratoire d'informatique, Recherche opérationnelle et Statistique Appliquée LIROSA.
Abdelmalek Essaadi, Tétouan, Morocco
soufiane.sail@gmail.com; halima.bouden@gmail.com

ABSTRACT

Nowadays the Cloud has started to gain ground even in SMEs, in spite of that the Cloud is still unknown for several ... for others few reliable.

SaaS represents a promising technology, which grows each year rapidly. Only at the security level, there are many obstacles, and becomes a major problem against its adoption. For example the public cloud represents a huge risk since the data of several companies are stored at the same place, close to each other.

For this reason the security of transiting or stored data in the SaaS remains a challenge for providers in order to gain the confidence of the customers.

SAAS remains the target of several attacks, such as network attacks, etc., which aim to disrupt its operation. Therefore, it is essential to deal with these attacks, and tried to minimize vulnerability and adopt new security concepts.

Through this document, we are going to study the security of SaaS, We will try to find the burly and feeble points of the most famous clouds such as Google Amazon, Microsoft, and come out with Countermeasures and proposals solution.

Keywords: Cloud Computing, SaaS, Internet attacks, Data security, Vulnerability, Authentication.

1 Introduction

Cloud computing, or simply the cloud, is not a fashionable effect as we hear from time to time in the history of computing. It is an innovation that will profoundly change our relationship with new technologies. Now computers, software and network equipment are no longer products that we buy.These are services that we rent, and we can separate from them when we no longer need them. [1] The cloud offer several services. SaaS represents one of the services of this technology, where companies can use applications, without the need to buy them or worry about installation, operating system compatibility etc. A sector that continues to grow with a high level service and competition has begun to play in favor of the customer.In spite of these advantages, the security of the SaaS presents a great problem for the companies, on the level of Confidentiality, integrity, authentication, location of data etc.This represents amajor obstacle for this technology.This reflects the lack of user confidence in Software as a Service. In this document we will study the security of SaaS, and then in the second part we

DOI: 10.14738/tmlai.54.3194
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3194

will do a comparison between Google Cloud, Amazon and Microsoft in terms of security, and finally see solutions that can be useful in securing the Cloud.

2 SAAS Security

As SaaS technology is one of the most widely used concepts in Cloud technology, it became the target of many attacks.Several security aspects in the SaaS are similar to the Web Service, since this technology requires only a web browser and internet in order to take full advantage of these services.In general, users do not dare to adopt such technology for fear of Attacks linked mainly to data security and more precisely to their confidentiality.

The most common problems with data security are data backup, data access, data storage locations, availability, authentication, and so on[2]. In the following table we summarize the different problems and the attacks of the cloud [Table 1].

2.1 Denial of Service (DoS) Attacks :

DoS is an attack that aims to make a service or resources unavailable to users for an indefinite period of time by flooding it with unnecessary packages.The main purpose of these attacks is to exhaust computer resources (CPU, network bandwidth) so that it renders services unavailable to users,The SaaS providers provided several services, if these services become inaccessible the customer will not be satisfiedwhich can have dangerous impacts. In a DDoS attack in general, the attacker typically disguises or "spoofs" the IP address section of a packet header in order to hide their identity from their victim.This makes it extremely difficult to track the source of the attack. [3]The DoS attack can use HTTP and XML to take down a server, this type of attack is called HX-Dos.

Table 1:Different problems and attacks on the Cloud

Security Issues	Attack vectors	Attack types	Impacts
Virtualization level Security Issues	• Social engineering • Storage vulnerabilities • Datacentre vulnerabilities and Network • VM vulnerabilities, etc.	• DoS and DDos. • VM Escape. • HypervisorRootkit.	• Software interruption and modification (deletion) • Programming flaws
Application level Security Issues	• Session management and broken authentication • Security misconfiguration, etc.	• SQL injection attacks • Cross Site scripting and • Other application based attacks.	• Modification of data at rest and in transit • Confidentiality • Session hijacking
Network level Security Issues.	• Firewall misconfiguration, etc.	• DNS attacks • Sniffer attacks • Issues of reuse IP address • Network Sniffing, VoIP related attacks (e.g. VoIP phishing).	• Traffic flow analysis • Exposure in network
Physical level Security Issues	• Loss of Power and environmental control	• Phishing Attacks • Malware injection attack	• Limited access to data canters • Hardware modification and theft

2.2 Authentification issues

The primary purpose of authentication is to give access to the right users. Authentication is a weak point in Cloud services, and it is often targeted by attackers. Today, most user-centric services still use a simple authentication username and password, with the exception of some financial institutions that have deployed various forms of secondary authentication Site, virtual keyboards, shared secret questions, etc. to make it a bit more difficult for popular phishing attacks etc. [4]

The main authentication attacks include:

2.2.1 Man in the middle :

This attack is performed when an attacker places himself between two users. Whenever attackers can place themselves in the path of communication between two or more people, they may intercept and alter communications. [4]

2.2.2 Key logger :

A key logger sometimes called a keystroke logger, a key recorder, or a system monitor, is a hardware device or a small program that monitors each keystroke that a user types on the keyboard of a specific computer. As the user types, the device collect each keystroke and save it as text in its own miniature hard drive. [5]

2.2.3 Phishing / Ingénierie sociale :

This is an attack that aims to convince a user to reveal sensitive information (password or other sensitive information). Internet users often suffer from this attack, and since the cloud is web based, Cloud Clients are concerned as well. [2]

2.3 SQL Injection and Cross-Site Scripting :

The two most common threats that are used to steal user information from the Web application are sql-injection and cross-site scripting (XSS) by inserting malicious code into the Web application. Therefore, when users return data by using the text box on a Web page, hackers add special characters. The malicious code is inserted into a standard SQL code that alters the natures of the query.The attackers access a database and execute their own SQL command. That can be used to drill, modify, and delete the information. [6] Cross-site Scripting is one of the most common application layer hacking techniques that inject malicious scripts into web content. As the Cloud provides a shared environment, attackers attempt to insert malicious scripts such as JavaScript html and VBScriptinto the dynamic web application as a browser-side script to gather important information. [7]

2.4 Password reset attack :

In such an attack, the hacker attempts any possible combination of characters in order to find the password. Strong encryption can make data difficult to decipher but not indecipherable, since exploitation of large resources and techniques can make it possible.

2.4.1 Brute Force Attack :

In general users choose words or phrases easy as password, this facilitates the attack. Based on a list of word and phrase the attacker tries to guessed and found the right password.

2.4.2 Brute Forcing Session Identifiers :

In this attack, the hacker attempts to guess the session ID of the targeted victim, in order to succeed this attack the hacker needs a very powerful computer, which can be possible, if a hacker rents or buys a high-performance machine At Amazon for example.

2.5 Side Channel Attacks :

An attacker could attempt to compromise the cloud by placing a malicious virtual machine in the immediate vicinity of a target cloud server and then launching a lateral channel attack. Lateral channel attacks have emerged as a kind of an effective security threat targeting the implementation of cryptographic algorithm systems. The evaluation of the rigidity of a cryptographic system to Side channel attacks is therefore important for the secure concept of the systems. [8]

3 Google cloud vs Amazon vs Microsoft

Google nowadays is one of the world's leading in the cloud, offering several services, Google Drive, Dropbox (photos...), Gmail (emails) etc. This giant has gained the confidence of millions of users around the world, with its innovations, quality and especially security.

Google Drive is one of Google's services, oriented data storage remotely. It is used to synchronize between a PC and the Cloud. [9] Google Drive is not mountable as a network drive, is not accessible via FTP or another open protocol, which limits network attacks of this kind.

But Concerns about Google Drive are usually related to legitimate security. Of course, like any cloud-based storage system, users are under the mercy of servers. The worst nightmare that can happen to a user is seeing the servers breaking down (Google is not immune to this, such a breakdown has already been reported before), which will cost a lot for Google and its Loyal customers. This is why reliability is an important issue.

Except that at this level Google has a rather good technical infrastructure, since it has expertise in this area that few other companies can match, as to the availability and speed of servers are almost exemplary.

One of the security measures adopted by this company to limit human error, which can be committed by the user, retention of deleted files for 30 days, which can be very useful,

The data is stored and fragmented, with a randomly chosen file name and encrypted.In spite of the progress it has made, it remains often under criticism, as only the data transiting and stored in the servers are encrypted, since Google does not offer the possibility of encrypting locally via a private key and chosen by the user, So Google is able to decrypt without user intervention. To catch up with competitors Google has just launched an encryption key management service (KMS) after a delay of two years in front of Amazon, this tool provides developers with a solution to manage encryption keys.

Recently a Google search team had tried to develop their own encryption algorithm, but there is no evidence that it will be adopted, at least in the near future. [10]

The security side then represents a greater brake at Google. For example, in order to connect to the storage space, we can use our Google account, gathering a username and password for two different services can cause serious and catastrophic damage, since if the user leaves Gmail account for example opened or the window of its browser without disconnecting, anyone can easily access and see all the files

stored. Google's advice regarding this topic remains basics, they basically consist of using a strong password, and logging out when finished etc.

Finally, this American giant, where the majority of its servers found it on American soil, must satisfy the requests of access of the American authorities according to the terms of Patriot Act; it means that they must communicate the information which is entrusted to them to the United States intelligence agencies. This puts users' privacy at risk. Google goes beyond that, since it states in their terms of use that they are entitled to use user's data, for a service improvement, a deliberately vague definition, which does not show the limits of use of those information.

On the other side of the competition, Amazon, another giant that has made fascinating progress, by dominating this market far ahead of Google and Oracle. It has achieved spectacular revenues. But Amazon is not immune from criticism, especially when it concerns the security of its cloud; with a high number of breakdowns compared to others. Protection systems have proved several times their inability to cope Attacks. Despite that progress has been done in order to improve their services and minimize vulnerabilities.

The Amazon Cloud is the source on which several services are based, so the reliability of the Amazon infrastructure is no longer to be demonstrated. The performance of Amazon Cloud Drive is also excellent.

However, Amazon is one of the worst providers in terms of the confidentiality of the data entrusted to it. The data is not encrypted locally and is stored in clear text on the servers. Encryption occurs only to protect data during transfer. Fortunately, there are third-party services offering to encrypt data before sending to Amazon Cloud Drive. [9]

Often Amazon is blamed for dealing with the security of these infrastructures more than web applications, causing flaws related to attacks like XSS and SQL Injection, to catch up, the company has tried to put up a layer of Security that will limit such fails, except that it did not give the desired results, since not so long ago Hackers were able to recover several password users, a blow that brought the company to adopt a method of double identification. Today plus your email and password Amazon will ask you to enter a special code (sent by SMS) in order to better secure accounts, a procedure that can protect more.

Amazon had also removed the local applications that allowed the connection to the data, and it limited itself from the website, meaning that to access its storage space the user is obliged to pass through the website.

Moreover, since Amazon is a company under American law, hosting on American soil, the data are also submitted to the Patriot Act, as for its competitor Google.

Finally with regard to Microsoft and its One Drive, it does not do better than the others. It is true that the transiting data is protected via SSL, and we can opt for two-factor authentication, except that once on the server the data are stored in clear, only professionals with a contract One Drive Enterprise have the chance to see their data stored encrypt.

Worse than that Microsoft reserves the right to remove some content that is contrary to the terms of use of its service, which means that customer data are analyzed.

Microsoft does not escape the Patriot Act law same like Google and Amazon.

4 Proposition and Solution :

The following key security features should be carefully considered as an integral part of the SaaS application development and deployment processes FIG1:

- Data security.
- Network security.
- Location of data.
- Data Integrity.
- Segregation of data.
- Access to data.
- Authentication and authorization.
- Data confidentiality.
- Web application security.
- Availability.
- Backup.
- Identity management and connection process. [11]

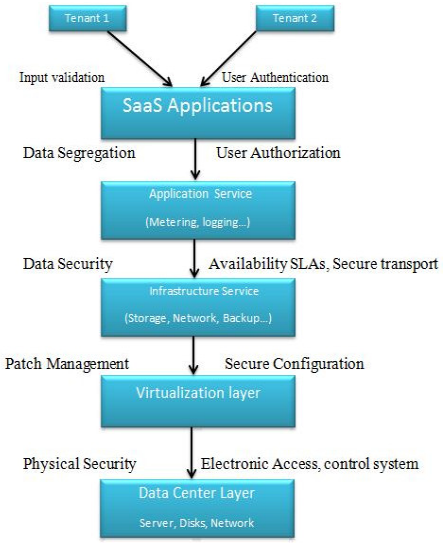


Figure 1:Different security layer in the SaaS

4.1 Network security and data integrity (transport):

SSL or Secure Socket Layer is a protocol to secure communication between two computers. This solution can solve problems related to the transport of information between a client and a server.

This system allows securing any protocol of the TCP / IP. Based on two protocols, that allows total security of the data transiting. If we use 128 bits - SSL we will have a total of 2 ^ 128 key combination possibility, which will be very difficult for a hacker to get to decrypt it, the SSL provided another version of 256bits even more secure. This one will be quite complicated for a hacker, and easy to deal with a brute force attack.

The notion of certificate (CA's Certificate Authorities) in the SSL during an SSL negotiation is considered as the guarantor, in order to clearly identify the server with which we communicate. The Cloud server first

sends the credentials to the client when it connects and then sends a copy of its SSL certificate. The owner verifies the certificate and then sends a message to the server and the server sends a digitally signed acknowledgment to start an encrypted SSL session to transfer the data between the browser and the server in encrypted mode. [12]

4.2 Data Storage :

The layered approach to secure data storage is given in FIG 2.

Where the first layer allows a secure authentication of the person, then the second layer will allow the anonymization of the person's data. Third step concerns the encryption of the data, so that finally they will be classified in order to well manipulate the data, and guarantee a high security.

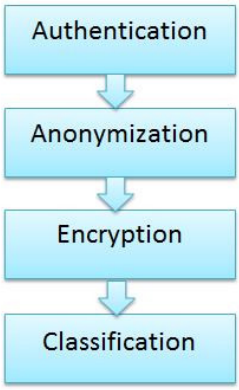


Figure 2: Proposed security model.

4.2.1 Authentication:

An effective average of authentication is OTP (one time password) which is a valid password only for a session or transaction. OTP is like a validation system that provides an extra layer of security. It is automatically generated by a pre-calculated method, which will eliminate some of the shortcomings associated to static passwords such as password longevity, password simplicity, and brute force attack. OTPs are generated on the server side and sent to the user using a telecommunication channel. [13]

Clouds cannot be accessed without the right combination of user name, user password, and one-time password. OTP generated must be difficult to estimate, retrieve, or trace by hackers. [14]

Several elements can be used to generate a single-use password that is difficult to guess. [13]

4.2.2 Encryption and Anonymization:

Since the data is no longer stored on the client side, and since the cloud is based on resource sharing technology, this sharing makes the exact location of user data impossible to determine. That does generate problem of security and accessibility. [16]

Encryption has become a necessity for the customer to protect his data. To increase the security level of data in the public cloud, it is important to encrypt them using anonymization with backups and audits. [17]Anonymization can be defined as the operation of deleting all the information that directly or indirectly identifies an individual [18] contained in a document or a database. This will make it very difficult to re-identify the persons or entities concerned.

4.2.3 Classification:

In order to better manipulate, and limit the leakage of the data, a method of classification of the data is adopted. This classification is generated by an algorithm that calculates the degree of sensitivity of data, based on 3 criteria (Confidentiality, Integrity, Availability) Fig3. Once the calculation is done, we classify our data according to 3 categories. Public, private, access limited. [19]

```
1. Input: Data, protection section, D [] array of n integer size.
   Where D[ ] array consisting of C, I, A, SR, R of n integer size.
2. Output: Categorized data for corresponding section.
3. For i=1 to n
        3.1 C [i]=Value of Confidentiality.
        3.2 I [i]=Value of Integrity.
        3.3 A [i] =Value of Availability.
        3.4 Calculate SR [ i ]=(C[ i ]+(1/A[ i])*10+I[ i ] )/2

4. For j=1 to 10
        For i=1 to n
        IF SR [ i ]== 1||2||3 then
        S[ i ]=3
        IF SR[ i ]== 4||5||6 then
        S[ i ] =2
        IF SR [i]== 8||9||10 then
        S[ i ] =1
```

Figure 3: Classification algorithm.

The result of our calculation will allow us to classify our data, a result Sr = 1 || 2 || 3 will classify our data in the category (high sensitivity) where access is limited, if Sr = 4 || 5 || 6 data will be placed in the private (private) category, otherwise the data will be stored in the public part.

This classification makes it possible to better manipulate our data and to reinforce security. Fig4

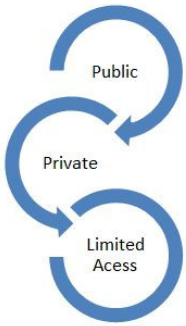


Figure 4: The different levels of classificatio

4.3 Location of data and Confidentiality:

In the Cloud and SaaS, users do not know where their data is stored; problem arises, when these data are stored in countries where laws allow governmental or other organizations to access these data. For example, the United States. To solve this problem, the service provider must immigrate to a country where the laws do not permit any private data access. A secure SaaS model must be able to provide reliability to the client on the location of the data.

4.4 Availability:

Service availability is one of the strengths of the cloud service, but at the security level, it is one of the biggest problems, the unavailability of a service for some time, can cost a lot for a customer. This customer

whowant to have a service available at any time. The DoSs attack is an example of attacks that aims to make a service or resource unavailable to users by sending useless messages.

In order to identify these messages, we propose the SOTA method which is based on the DPM algorithm [21] [22] (DPM's role is to mark the ID field and the reserved FLAG in the IP header). SOTA, which is a web service security application, its objective is to apply the traceability of SOA approach. In order to identify the false identities of the messages, since the objective of the X-DoS, and DX-DoS and hidden the true identity of the attacker.

Each packet entering the router will be marked. The marked packet will not change during its passage through the network. The DPM is applied to the SOTA Framework, placing the Service Oriented Trace back Mark (SOTM) mark in the web service message. The true identity of a package is stored in the SOTM and placed inside the SOAP (web service). Like DPM, in SOTM the tags do not change during the transit in the network. SOTM is composed of an XML tag and placed in the SOAP header. When discovering an X-DoS or DX-DoS attack, SOTM can be used to identify the true source of fake messages.SOTA only offers the ability to identify fake messages that aims to disrupt a server by flooding it with unnecessary messages to exhaust it. The direct elimination of a message from an X-DoS and DX-Dos attack will be supported by the Cloud Protector. [20] Which is a Neural Network (Neural Network), who allows X-DoS messages to be detected and filtered. A neural network is a set of connected units composed of input and output layers. In a neural network, the threshold logic unit (TLU) is used. The TLU inserts the input objects into an array of quantities and additions to each other, to see if they are above fixed threshold or not.

Other solutions proposed to limit the impact of network attacks (Dos, DDos etc.). The IDS which is a set of software and hardware components that primarily aims to detect and analyze any attack attempt. NIDS analyzes the network to detect attacks based on a comparison between the analysis and the standards. And HIDS which is more interested in the machine than the network. It analyzes in real time the flows related to a machine.

The combination of these two will have a positive effect, if they are properly configured.

The life cycle of a package in such a model is simple:

Once the packet arrives at the firewall it will be analyzed based on the configuration, if it is invalid it will be blocked otherwise it will be processed by our IDS (NIDS & HIDS) Based on defined standards and configurations. If the system will recognize the packet as innocent it will redirect to the requested Cloud service, otherwise the system will trigger the alarm and it will notify the cloud administrator. All the packets received from same IP will be deleted, and the system will immigrate to another more secure virtual machine, and we will inform the firewall of the new configurations that it must take into consideration. [23]

4.5 Application security:

In security, humans remain the weak point; some mistakes committed can be fatal. Those errors that can be formatted in bad configuration as they may be related to programming the application and codes.Such errors are catastrophic, since they open the doors to hackers to harm the system; SQL Injection XSS etc. are all attacks that are based on flaws in a code.This is why application security tests are very important before commissioning.

Web Application Security Scanner is software that performs automatic tests on an application and identifies security vulnerabilities. The scanners do not act on source code, but they only perform functional tests and try to find security vulnerabilities. We can find several types of analyzer, whether open source or paid, for example Grabber, Wapiti, Zed Attack Proxy etc. [24]

5 Conclusion

As already mentioned in this document, the SaaS and the Cloud in general offers great advantages for companies, at the technologic level, cost, flexibility and ease of use etc.

Only the obstacles that limit the adoption of this technology are numerous, and security comes at the top of these obstacles. Since some practices can seriously harm the operation of the Cloud. For example collecting all the data from the different clients in one location puts all data at risk.

The flexibility and ease of adapting SaaS to data access policies in an enterprise is paramount, it will protect data, and ensure the integrity and confidentiality of data.

A SaaS must be reliable, to convince a user, whose only concern is whether his data is secure or not.

In this document, we tried to deal with some of the attacks that target SaaS, offering some solutions that can ensure data security, such as SSL which is one of the best solutions to protect the circulating data, CTB which allows Identifying incoming packets to protect against DoS attacks, or the data encryption and classification model, which will help secure the stored data.

REFERENCE

[1] Emile Yaogo« Le cloudcomputing : L'informatique comme l'électricité ou l'eau que nous consommons ».http://www.faso-tic.net/spip.php?article415&rubrique1 29décembre 2016.

[2] Salman Iqbal, Miss Laiha Mat Kiah, BabakDhaghighi, MuzammilHussain, Suleman khan, Muhammad Khurram Khan, Kim-Kwang Raymond Choo. On Cloud Security Attacks: A taxonomy and Intrusion Detection and Prevention as a Service. 2016

[3] I. Mettildha Mary, P.V.Kavitha, Priyadharshini M, Vigneshwer S Ramana. Secure Cloud ComputingEnvironmentagainst DDOS and EDOS Attacks. 2014

[4] Ajey Singh, Dr. ManeeshShrivastava. Overview of Attacks on Cloud Computing. 2014

[5] Margaret Rouse. http://searchmidmarketsecurity.techtarget.com/definition/keylogger

[6] Bhadauria, R., Al., A survey on security issues in cloud computing. 2011

[7] Rodero-Merino, L. et Am., Building safe Pass clouds: A survey on security in the multitenant software platforms. Computer & security, 2012.31(1): p 96-108

[8] Qiasi Luo1 and Yunsi Fei2. Algorithmic Collision Analysis for Evaluating Cryptographic System and Side-Channel Attacks", International Symposium on H/w – Oriented Security and Trust, 2011

[9] Matthieu Lamelot, Stockage en ligne : lequel choisir ? , http://www.tomshardware.fr/articles/comparatif-stockage-cloud,2-2332.html 7 Novembre 2016

[10] VALENTIN BLANCHOT Une IA de Google a créé son propre cryptage de données. https://siecledigital.fr/2016/11/03/intelligence-artificielle-google-cryptage-donnees/ 3 Novembre 2016.

[11] S. Subashini, V. Kavitha, A Survey on Security Issues in service delivery models of cloud computing. 2011

[12] Sandeep K. Sood, A combined approach to ensure data security in cloud computing.

[13] : Al Haddad Zayed, Hanoune Mostafa, MamouniAbdelazize, « Cloud Computing et sécurité : Approches et solutions. Décembre 2015.

[14] Balakrishnan.S, Saranya.G, Shobana.S, and Karthikeyan.S, « Introducing Effecyive Third Party Auditing (TPA) for Data Storage Security in cloud » Int. J.Comput.SciEnceTechnol, vol. 2 Jun 2011.

[15] F. Aloul, S. Zahidi, and W. El-Hajj, « Two factor authentification using mobile phones, » in Computer Systems and Applications, 2009 AICCSA 2009. IEEE/ACS InternationalConference on, 2009, pp. 641-644.

[16] E. M. Mohamed, H.S Abdelkader, and El-Etriby, « Enhaced data security model for cloud computing, » in Informatics and Systems (INFOS), 2012 8th International Conference on, 2012, pp. CC-12.

[17] Insitute of Electrical abdEkectronics Engineers, Ed., « Enhancing Data Security during Transit in Public Cloud , » Int.J. Eng. Innov. Technol. IJEIT, vol.3 ,Jul. 2013.

[18] « la protection des données personnelles dans l'open data : une exigence et une opportunité. » [online]. Available :http://www.senat.fr/rap/r13-469/r13-4697.html. [Accessed : 08-jul-2015].

[19] SagarTirdkar, YazadBaldawala, SagarUlane, Ashok Jori. Improved 3-Dimensional Security in Cloud Computing. International Journal of Computer Trends and Technology (IJCTT°- volume 9 number 5. 2014.

[20] AmanSagar, Bineet Kumar Joshi and NishantMathur. A study of distributed denial of service attack in cloud computing (DDOS). 2013.

[21] Belenky A, Ansari N. Tracing multiple attackers with deterministic packet marking (DPM). In : Proceedings of IEEE Pacific Rim conference on communications, computers and signal processing, vol , 2003 p.49-52.

[22] Ashley Chonka, Yang Xiang, Wanlei Zhou, AlessioBonti. Cloud Security defence to protect cloud computing against HTTP-DOS and XML-DOS attacks. 2010

[23] Omar Achbarou, My Ahmed El kiram, Salim El Bouanani, Securing Cloud Computing from different Attaks Using Intrusion Detection Systems.

[24] Ashley Chonka, Yang Xiang, Wanlei Zhou, AlessioBonti. Cloud Security defence to protect cloud computing against HTTP-DOS and XML-DOS attacks. 2010

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Electrical Behavior of Solar Cell Based on ZnO/PS

[1]Z.Yamlahi Alami, [1]J.El khamkhami, [2]M. Salem and [2]M.Gaidi
[1]laboratoire de la matière condensée, dept .de physique, faculté de sciences Avenue de Sebta, Mhannech I 93002 - Tétouan – Maroc
[2]Laboratoire de Photovoltaïque Centre de Recherches et des Technologies de l'Energie, Technopole de Borj-Cédria, BP 95, 2050 Hammam-Lif, Tunisia
a.yamlahizineb@gmail.com

ABSTRACT

Because of high loss of photovoltaic conversion due to reflection of incident photon by the silicon surface, we proposed in this work a single heterojunction solar cell model based on porous silicon (PS) and Zinc oxide to minimize this loss, which the ZnO could act like front n-layer as well as antireflection coating saving processing cost and complexity. In the experiment the ZnO layer was deposited by spray pyrolysis method on porous silicon (PS) using nitrate precursors with different molar concentration. The proposed treatment has also demonstrated a significant decrease in the reflectivity of ZnO/PS. To simulate a solar cell based on porous silicon and zinc oxide, we have used PC1D software. The variation of internal and external quantum efficiency, and I-V characteristics have been done by changing the reflectivity of ARC layer, The I-V results obtained by PC1D simulation was compared with the experiment results: Simulations anticipated conversion efficiency was 14,5% while by experiments was 12% for the device fabricated with 0.25 M.

Keywords-component; ZnO, PC1D, Characteristic I-V.

1 Introduction

The studies have focused on solar cells as an alternative energy sources to decrease the fossil fuels 'uses, but the high price of 1KWh of electricity generated by a solar cell prevents the wide propagation of this technology so the researches have concentrated on reducing this cost by using a cheaper materials as a cover layer .

Thin film nanostructure has unique proprieties make it capable to play a role as a cover layer. Zinc oxide (ZnO) is one of the most extensively studied materials as an alternative photo anode to TiO2 [1]. ZnO has a wide range of properties that depend on doping, including conductivity which may change at will from metallic to insulting (including n-type and p-type conductivity), high transparency [2], Piezoelectricity [3], room-temperature ferromagnetism [4], and huge magneto-optic effect [5] ,etc…

Without much effort, ZnO can be grown into many different nanoscale forms, thus allowing various novel devices to be achieved. It has a wide band gap energy of 3.37 eV, the larger exciton binding energy of 60 meV guarantees efficient luminescent and photovoltaic characteristics , a wide range resistivity, high electron Hall mobility (200 cm2 .V .s−1) , low toxicity , natural abundance , and highly stable wurtzite structure .

DOI: 10.14738/tmlai.54.3195
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3195

ZnO thin films have been prepared using various methods as magnetron sputtering [6], spray plasma [7], sol gel spin coating[8], pulsed laser deposition[9], metal-organic chemical vapor deposition and spray pyrolysis[2].

Among this method spray pyrolysis is a versatile technique for the deposition of metal oxides because of its low cost and easy process control which gives the possibility of obtaining films with the required properties for different applications [10].

In this paper, we propose a new model of potentially cost effective and high efficiency single heterojunction solar cells based on porous silicon (PS) (rear region) and the emerging II–VI material ZnO(front region).The ZnO grown on porous silicon can work as an active n-layer as well as antireflection coating.

This model was simulated by PC1D software, the different electrical and optical characteristics as internal, external quantum and efficiencies are investigated by varying different key parameters.

2 Experimental

Experiments were carried out on p-type, boron-doped multicrystalline silicon substrate with a thickness of 350µm and a resistivity of 0.5 to 2 Ω cm. Porous silicon was formed by electrochemical etching mc-Si wafers in HF 48% /ethanol (1V:2V) at a constant current density of 15 mA/cm2 for an etching time of 10 min. Immediately after PS formation, the samples were rinsed with DI water and dried under N2 flux[10].

ZnO thin films were prepared by spray pyrolysis process from a solution with different molar concentrations of Zn(NO3)2,6H2O precursor (0.05 M, 0.1M, 0.15 M , 0.2M and 0.25M) in 25 ml of deionized water. The spray system used in this work was fully described and schematically presented in [11]. The Zn (NO3)2,6H2O solution was sprayed with a solution spray rates of 2 ml/min onto a preheated silicon porous substrate at 500°C during 10 min, using compressed air as a carrier gas. The nozzle to substrate distance was about 45 cm. The ZnO thin Films thus obtained are transparent, exhibiting a good adherence to the substrate surfaces.

The reflectivity spectra of the different surface were measured using a LAMBDA 950 UV/ Vis/NIR spectrophotometer equipped with an integrating sphere. The effective minority lifetime (τeff) was measured using the Sinton WCT-120 lifetime tester in quasi-steady-state photoconductance (QSSPC) mode.

3 Simulation

Our experiment results were introduced in the PC1D simulation software to simulate the solar cell based on porous silicon and ZnO using the parameters illustrate in table 1.

Table 1 resumes the parameters was optimized for the solar cell based on the some values obtained in the experiments characterization and the other based on the ZnO proprieties.

The several adjustable parameters in the PC1D can be iterated, to introduce the EQE, IQE, I-V and efficiency of solar cell.

Table 1: Optimized Parameters for the Solar Cell used by PC1D

Varying parameters	Value
Device parameters	
Device area	1 cm^2
Surface texturing	None
Surface charge	None
Front reflectance	18%-9%
Emitter contact resistance	$10^{-6}\Omega$
Base contact resistance	0.4 Ω
Front region (ZnO)	
Region thickness	0.015µm
Electron/hole mobility	50 cm^2/Vs
Dielectric constant	8.66
Bandgap	3,64- 3,14 eV
Refractive index	2-2,38
n-type background doping	3.10^{13}
bulk electron/hole recombination time	1,2 -1,6 -5- 57-90µs
electron/hole front surface recombination velocity	1.107 cm/s
Rear region (Si)	
Region thickness	350 µm
Carrier mobilities	From internal model of PC1D
Dielectric constant	11,9
Bandgap	1,124eV
Intrinsic concentration at 300 K	1.1010 cm^{-3}
p-type background doping	1.1016 cm^{-3}
peak rear p-type doping	3.1018 cm^{-3}
bulk electron/hole recombination time	100µs
electron/hole front surface recombination velocity	1.1016 cm/s
electron/hole rear surface recombination velocity	300 cm/s

4 Results

4.1 Experimental results

Some parameters were changed when we have changed the molar concentration of the Zn, these parameters have influenced on the simulated results as EQE, IQE, I-V, and solar cell's efficiency, where the reflectivity parameter of the thin film ZNO/PS is the most important. Lower is the reflection from the front surface higher is the EQE and IQE.

Figure.1 Shows AFM images of the ZnO/PS surface treatment and a reference sample (untreated PS layer) for comparison. It can be seen from the different AFM topography a clear evolution of the microstructure and the surface has an important roughness, the reported roughness (RMS) was found to be around 425 nm for PS wafer and increases to 556 nm after ZnO deposition. The RMS enhancement is attributed to the increase of the nanoparticle size when increasing ZnO molar concentration.

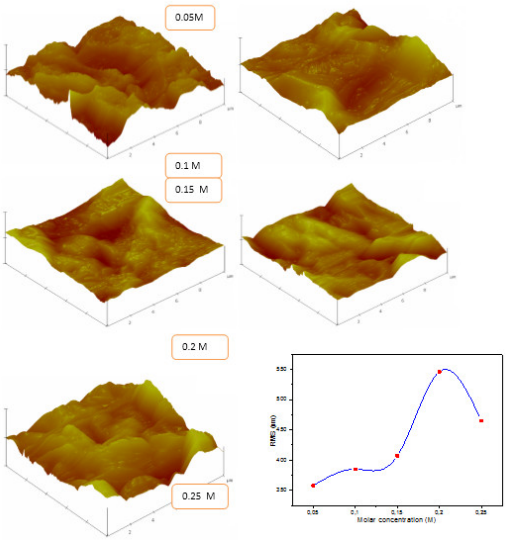


Figure.1 (3D) AFM images of PS and ZnO/PS with various Zn concentrations

Figure. 2 shows that the average reflectivity reduced from 18 % to 9% when the molar concentration decreases from 0.25M to 0.05M. The measured refractive index and extinction coefficient values changed, the former from 2 to 2.38 which is very close to the ideal value and the later from 0.01 to 0.04 for λ = 500 nm [10]. This can eliminate requirement of additional antireflection coating, resulting in reduced fabrication process steps and cost. For Babar et al. [14] the refractive index of ZnO at 600 nm equal 2 is very suitable to play role of antireflection coating AR for Si. The effective carrier lifetime also another important parameter influences on the simulated result (figure 3 shows the values of the each sample obtained on the experimental results).

The different experiment results obtained have introduced using PC1D to simulate the different parameters as EQE, IQE, I-V, and solar cell's efficiency.

B.Hassain in his work [14] also observes that the VOC , Jsc and efficiency (η) have influence by change of ZnO thickness using PC1D software , where the optimized modeled thickness was obtained as 0.1µm for minimum reflection, they have achieved an efficiency of 19,3% , a Voc of 622 mV , a Jsc of 38,5 mA .

4.2 Simulated results

We can distinguish between two types of quantum efficiency: external and internal (EQE and IQE), in other words the external quantum efficiency is defined as the ratio of the charge carrier's number collected by the solar cell to the photon's number of a given energy incident on the solar cell from outside. The second kind (IQE) is the ratio of the charge carrier's number collected by the solar cell to the number of photons of a given energy that shine on the solar cell from outside and are absorbed by the cell. A low value of thin film's IQE indicates that the active layer of the solar cell is unable to make good use of the photons.

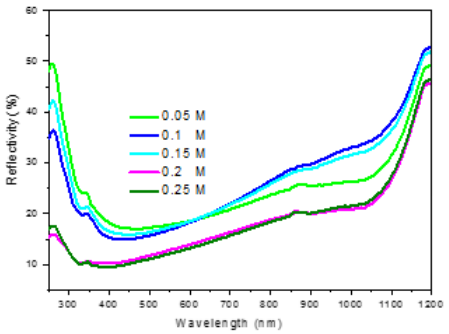


Figure.2 Total reflectivity ZnO/PS film

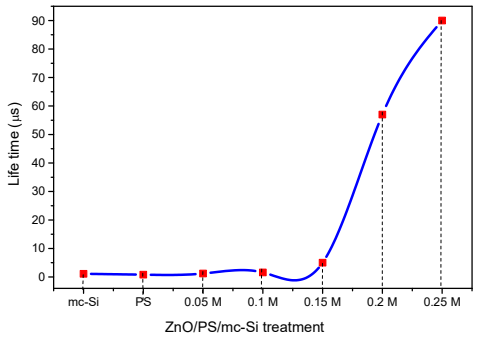


Figure 3: Effective minority carrier lifetime evolution of ZnO/PS

Figure 4 shows the variation of IQE with different front back. The graph indicates that the IQE augments with the increase in the molar concentration of ZnO thin film due to the low value of reflection. The low reflectivity of the thin film ZnO/SP fabricated with a high concentration of Zn (0.25M and 0.2M) also influences on the wide of the band. So the thin film with a high molar concentration has absorbed the photons with long wavelength. In a similar way EQE also increases with the increasing of molar concentration (these results are shown in Figure 5). The absorption of high energy photons in ZnO is mainly responsible for poor EQE at short wavelengths.

Table 2 The Eg values of different n-ZnO

Sample molar concentration (M)	0.05	0.1	0.15	0.2	0.25
Eg (eV)	3.64	3.56	3.34	3.15	3.15

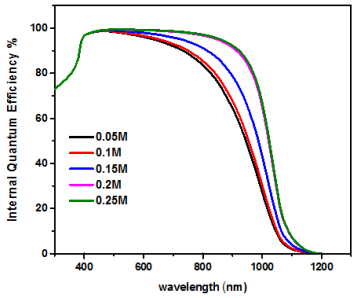


Figure. 4 Internal quantum efficiency (IQE) of the solar cell device with different front back.

The EQE was clearly improved in which the values of EQE increased from 80 to 90% at a wavelength of 500 nm .This result might be explained by the decreased reflectivity of ZnO/SP. Babar has explained in his work [13] the improvement of EQE and front surface reflection with taking into account antireflection effect of ZnO layer, therefore, the EQE was increasing from 84% to 98% around wavelength of 600nm when he adds the ZnO layer.

When the Eg decreases, the wideness of band gets larger, therefore, the IQE and EQE got influenced. All this due to fact the solar cell could absorb more the photons characterized with lower energy. Thus, the number of photons absorbed increases, (table 2 shows the value of the Eg of each sample).The decrease in band gap energy with increase in molar concentration could be due to increase in crystalline size and reduction of defect sites and increase in film thickness [2, 10] .

Figure 6 shows the comparison between the experimental I-V curve and the simulated one. The slight increase in the I-V simulated curve due to the front contact of solar cell is ideal in the simulation.

Based on life time values obtained in the experiment and reflectivity data, we have stimulated by PC1D the I-V spectra of ZnO /SP heterojunction under AM1.5 illumination.

The simulated I-V curves of the HJ solar cell of ZnO/SP treated with different molar concentration are shown in figure 7. The enhancement in the value of the short-circuit current density (JSc) from 26,8 mA to 34, 77 mA and the open circuit voltage Voc from 526,1mV to 536,9 mV .The Jsc increase monotonically with increased Zn molar concentration because of significant increase in number available photons in space charge region as shown in figure 4.

The I-V curves simulated shows the dependency between I-V behavior and different parameters obtained in the experience those parameters can be cited as follows the reflectivity, life time and band gap of ZnO thin film deposited on Si/porous (table 1 and table 2) , optimizing these three parameters yields an optimum Jsc and Voc .

Salem et al. in his work [10] has demonstrated an improvement in the short-circuit current density (Jsc) from 23.4 to 30 mA/cm3 ,and the open circuit voltage Voc got increased from 526.1 mV to 536.9 mV, in which these values are close to the result obtained in this work.

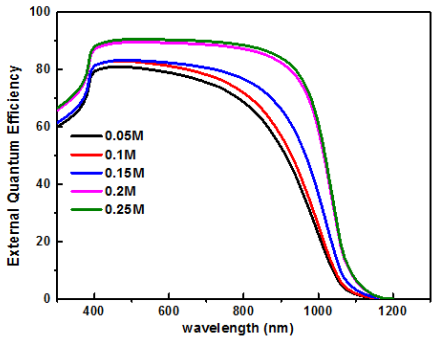


Figure.5 External quantum efficiency (EQE) of the solar cell device with different front back

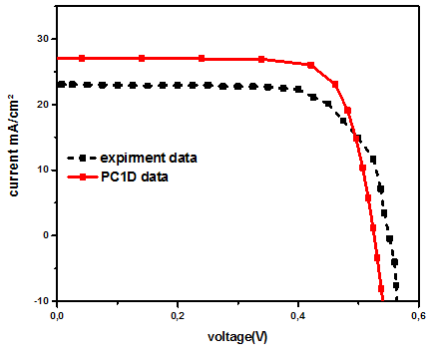


Figure . 6 I-V Experiment and PC1D spectra of ZnO /PS of 0.05M sample

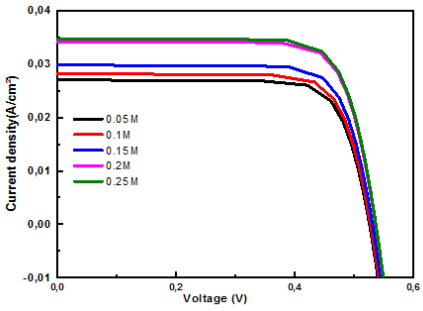


Figure .7 I-V Characteristics of ZnO / PS heterojunction under AM1.5 M

The quality of the solar cell is defined by it efficiency (η) of light conversion; the efficiency is the ratio of the energy output from the cell (Pmax) to the input energy from the sun (Pin). It was calculated by using the equation:

$$\eta = \frac{\mathrm{Im} * Vm}{Pi} \quad (1)$$

Where Vm is the value of the voltage for the maximum power from a solar cell, and Im is the value of the current for the maximum power from a solar cell.

The η values go up from 11.1% to 14.5%. The proposed treatment demonstrates a significant enhancement of this parameter after ZnO/PS treatment.

On the other hand the result of I-V and η in the experiment results are shown in the table 3, as it is seen there is the slight enhancement in the simulation results due to the ideal solar cell.

Table 3 Photovoltaic parameters from illuminated characteristics of solar cells obtained by PC1D software.

samples	Jsc(mA/cm3)	Voc(mV)	η(%)
0.05M	26,8	526,1	11,1
0.1M	28 ,19	527,8	11,6
0.15M	29,86	531,1	12,3
0.2M	34,13	536,4	14,3
0.25	34,77	536,9	14,5

Table 4 Photovoltaic parameters from illuminated (I–V) characteristics of obtained solar cells by experiment [10]

samples	Jsc(mA/cm3)	Voc(mV)	η(%)
0.05M	23,4	551	7 ,8
0.1M	23	552	9 ,1
0.15M	24	552	10 ,5
0.2M	28	553	10,9
0.25M	30	564	12

5 Conclusion

We have proposed a single hetero-junction solar cell model based on porous silicon and ZnO using modified PC1D software. ZnO has fabricated by spray pyrolysis from a solution of Zinc nitrate with different molar concentration. The influence of molar concentration full filled on the electrical and optical proprieties. We have obtained a low reflection with a high molar concentration of precursors, where the reflectivity of the device fabricated with 0.25 M is 9%. The effective carrier lifetime also increases in which the value of this device is 90µs. The different parameters introduced by PC1D software, the IQE, EQE, the I-V and the device efficiency value of this device have improved, in which the value of efficiency is 14.5% by PC1D software and 12% in the experiment.

REFERENCES

[1] Long Yang, Bao-gai Zhai , Qing-lan Ma, Yuan Ming Huang ," Effect of ZnO decoration on the photovoltaic performance of TiO2 based dye sensitized solar cells," Journal of Alloys and Compounds , pp 109–112 2014.

[2] Z.Yamlahi Alami,M. Salem, M. Gaidi, J.Elkhamkhami , "Effect of Zn concentration on structural and optical proprieties of ZnO thin films deposited by spray pyrolysis" Advanced Energy: An International Journal (AEIJ), Vol. 2, No. 4, 2015.

[3] Brijesh Kumar, Sang-Woo Kim,"Energy harvesting based on semiconducting piezoelectric ZnO nanostructures," Nano Energy ,vol 1, pp 342–355, 2012.

[4] Daqiang Gao, Zhaohui Zhang, Junli Fu, Yan Xu, Jing Qi and Desheng Xue, "Room temperature ferromagnetis of pure ZnO nanoparticles" J. Appl. Phys, vol 105, p 113928 ,2009 .

[5] A.J. Behan, J.R. Neal, R.M. Ibrahim, A. Mokhtari, M. Ziese, H.J. Blythe, A.M. Fox and G.A. Gehring, "Magneto-optical and transport studies of ZnO-based dilute magnetic semiconductors," Journal of Magnetism and Magnetic Materials , Volume 310, Issue 2, p. 2158-2160,2007

[6] C. Besleaga, G.E. Stan, A.C. Galca, L. Ion, S. Antohe, "Double layer structure of ZnO thin films deposited by RF-magnetron sputtering" Applied Surface Science 258 (2012) 8819– 8824.

[7] K. Bab, C. Lazzaroni, O. Brinza, M. Nikravech,"Effect of zinc nitrate concentration on structural and optical properties of ZnO thin filmsdeposited by Spray Plasma device," Thin Solid Films , vol 558 ,pp 62–66, 2014.

[8] Zi-Neng Ng, Kah-Yoong Chan, Thanaporn Tohsophon, "Effects of annealing temperature on ZnO and AZO films prepared by sol–gel technique"Applied Surface Science, vol 258 ,pp 9604– 9609 , 2012.

[9] L. Moreno, C. Sánchez-Aké, M. Bizarro, "Double-beam pulsed laser deposition for the growth "of Al-incorporated ZnO thin films"Applied Surface Science 302 (2014) 46–51.

[10] M. Salem ,Z. Yamlahi Alami ,B. Bessais , A. Chahboun , M. Gaidi, "Structural and optical properties of ZnO nanoparticles deposited on porous silicon for mc-Si passivation" J Nanopart Res .17 (2015) 137

[11] B. Elidrissi, M. Addou, M. Regragui, A. Bougrine, Lebanese Sci. Res. Vol 4, p 230, 1998

[12] M.Belarbi ,A.Benyoucef ,B.Benyoucef, " Simulation of the solar cells with PC1D, application to cells based on silicon ," Advanced Energy: An International Journal (AEIJ), Vol. 1, No. 3 ,2014 .

[13] Babar Hussain, Abasifreke Ebong, "Specifications of ZnO growth for heterostructure solar cell and PC1D based simulations", Data in Brief volume 5, pp 516–521, 2015.

[14] Babar Hussain, Abasifreke Ebong,Ian ferguson ,"zinc oxide as an active n-layer and antireflection coating for silicon based heterojunction solar cell ," solar energy materials and solar cells , vol 139 , pp 95-100,2015

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Comparative Study of Harris and Active Contour Using Viola-Jones Algorithm for Facial Landmarks Detection

Manal El Rhazi, Arsalane Zarghili, Aicha Majda
Intelligent Systems and Applications Laboratory, Faculty of Science and Technology
Sidi Mohamed Ben Abdellah University, Fez, Morocco
manal.elrhazi@usmba.ac.ma; arsalane.zarghili@usmba.ac.ma; aicha.majda@usmba.ac.ma

ABSTRACT

In this paper, we present a comparative study of two methods: Harris corners detector (H) and Active Contour detector (A.C) using Viola-Jones algorithm (V.J) for facial landmarks (eyes, nose, mouth) corners detection. These methods were implemented on two face databases; ECVP and FEI databases with combination of two methods (V.J + H) and (V.J + A.C). Experimental results showed that (V.J +A.C) gives a higher rate of detection than (V.J + H) method.

Keywords—Harris corner detector; Snake; facial features; Viola and Jones; image processing.

1 Introduction

The detection of facial features (eyes, nose and mouth) is the most important initial step in various facial image interpretation works related to computer vision application, such as: face identification, facial expression recognition, facial features tracking, facial beauty analysis…, and for more analysis accuracy, this detection includes the localization of facial features corners (eyes corners, mouth corners and nose corners).

The problem of features corners detection with accuracy and robustness has been a challenging task and has received a lot of attention in the past decade. So many approaches have been proposed in the literature to extract the facial features from an image and their corners; first for features localization, Viola and Jones method [1] has been widely used in many applications to detect the face and the facial features zones on an image. This method is the most stable and has shown a successful result in several applications, due to its basis which consists of many contributions: the cascade of classifiers trained by Adaboost [2]; for each stage in the cascade, a larger set of features is chosen based on the Adaboost. In addition, Viola and Jones method uses the integral image that helps in the speed feature evaluation [3]. Finally, the Viola and Jones is a powerful method to avoid the detection of regions that do not contain the interest object.

A large margin of approaches has been proposed recently for facial corners detection [4], for example, Harris corner detection [5] which is the most used in that field thanks to its facility of use and its good result of detection dealing with different face position and different level of luminosity. Another example is the active contours (or snakes) introduced by Kass and Witkin [6], they are curves that can deform progressively in order to be close to the outline of an object. This deformation is guided by minimizing an

DOI: 10.14738/tmlai.54.3196
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3196

energy function comprising two terms: Eint an internal energy which helps regulate the outline and an external energy Eext relied to the image and to specific constraints that can be added.

In the present work, we have used the active contour and the Harris corner detector combined with the facial axes that go along the features detected to improve the corners detection.

The purpose of this paper is to compare the rate of detection for both the Harris corner detector and the Active contour (Snake) using the Viola and Jones method, in order to detect the facial landmarks. The experiments were implemented on two face databases: FEI database [7] and ECVP database [8].

The global architecture for corners detection presented in this work is shown in Figure. 1

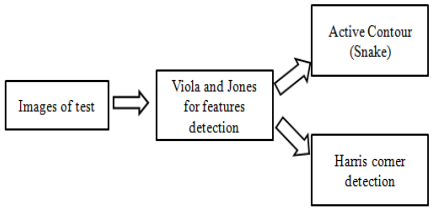


Figure 1. The global architecture for corners detection

This paper is organized as follow: section 2 explains the facial features detection using Viola and Jones, section 3 describes the features corners detection using the two approaches, section 4 will present the experimental results and section 5 contains the conclusion of this paper.

2 Facial features detection

2.1 Face detection

Before we detect the corners, face and facial features must be detected. In this paper we will use the Viola-Jones Adaboost detection method improved by R. Lienhart [9] to detect the face regions in global then eyes, nose and mouth.

Face detection is obtained by delimiting the area of interest by a rectangle based on Haar like features and cascade of classifiers learnt to detect a face zone on an image.

The figure below presents the rectangles using in the detection by the extended Haar like features:

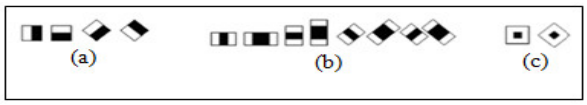


Figure 2. Extended Haar like features: (a) edge features, (b) line features, (c) center surround features

These features are used to calculate the difference between the sum of the white pixels areas and the sum of the black pixels areas.

$$d = \sum (white\text{-}pixels) - \sum (black\text{-}pixels) \quad (1)$$

To delimit the face region, a window of initial size of 24×24 (increased iteratively) scans the input image in all directions to find the face zone. This operation generates many features, which make the use of integral image required for the speed of detection. Then a cascade of classifiers based Adaboost is used

to classify the zone as a face or non face depending on the value of its descriptor obtained during the training using images of face and non face.

A cascade classifier consists of several simple classifiers which are applied one after the other on a region of interest in an image, while Boosting means to combine the results obtained by several "weak" classifiers to build one more efficient.

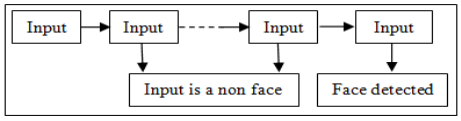


Figure 3. The cascade classifier

The integral image at the pixel (x,y) of an input image is calculated by summing the pixels above and left of the current pixel (x,y):

$$ii(x,y) = \sum_{x'\leq x,\ y'\leq y\leq}(x',y') \quad (2)$$

After this step, the algorithm of face detection return the face region localized (Fig. 4), which will be considered as inputs of the algorithm of facial features detection.

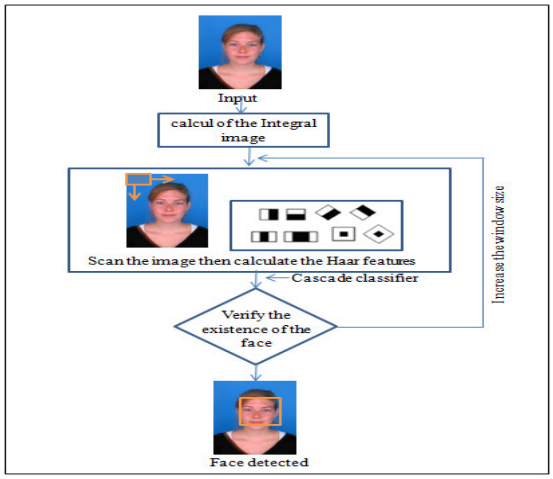


Figure 4. Face detection process

2.2 Facial features detection

To detect facial features, the result face detected in the previous step has been used with different extended Haar like classifier based Adaboost trained using different sets of images. The same steps have been followed for each facial feature (eyes, nose and mouth).

The result of face and features detection by Viola and Jones based Adaboost method, is given in the following figure for each database used.

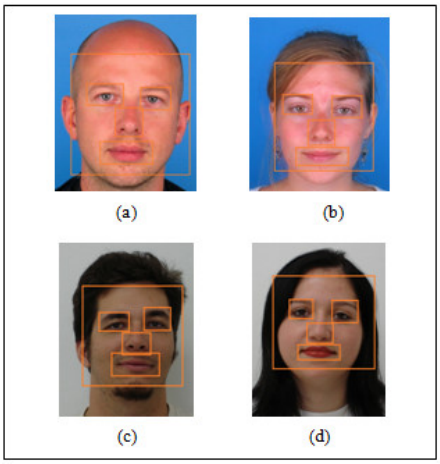


Figure 5. Facial features detection: (a) and (b) from ECVP database, (c) and (d) from FEI database.

3 Features corners detection

3.1 Harris corner detector

Harris [5] is the most popular detector for corners detection, it is based on the research of the zones in an image that has a change in intensity in multiple directions.

To capture the corners, we consider that the intensity change in direction (u,v) is given by:

$$change(u,v) = \sum_x \sum_y [(I(x+u, y+v) - I(x,y))^2] \quad (3)$$

The following matrix form describes the image I area at a given point:

$$E(u,v) = [u,v] M_{xy} [u,v]^t \quad (4)$$

With: $$M_{xy} = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (5)$$

And: $A = \frac{\delta I^2}{\delta x}$, $B = \frac{\delta I^2}{\delta y}$, $c = (\frac{\delta I}{\delta x}\frac{\delta I}{\delta y})$

$\frac{\delta I}{\delta x}$: X Sobel operator.

$\frac{\delta I}{\delta y}$: Y Sobel operator.

To determine the area's type, the eigenvalues are calculated, but since it is computationally expensive, an equivalent calculation of the term "R" is performed:

$$R=det(M) - k(trace(M))^2 \quad (6)$$

With: $det(M) = AB - C^2$ and $Trace(M) = A+B$ and typically k=0.04.

Depending on the value of R, the result can be interpreted in different ways:

- If R>0 then the area is a corner.
- If R<0 then the area is contour.

If R has a small value, then the area is a uniform zone.

For corners detection, we extract the local maxima and we keep just those with the highest value; which are the areas of corners.

3.2 Active contour (Snake)

An active contour [6] is a closed curve on an image, which can be iteratively deformed due to the external constraint forces and influence of the image forces and pull towards the contours of the segmented object in the image. This deformation is guided by the minimization of an energy function to achieve equilibrium.

This energy is composed of two terms: an internal energy that serve to impose a piecewise smoothness of the snake, and an external energy function that push the snake towards salient image features, and place the snake near to the desired local minimum.

If a snake is defined as a parametric curve **v**(*s*) = (x(*s*), y(*s*)), the energy function is given by:

$$E_{snake} = \int_0^1 E_{internal}\big(V(s)\big) + E_{external}\big(V(s)\big)ds. \quad (7)$$

The external energy can be divided on two energies; Eimage refers to image energy that attracts the snake towards desired features and Econst represents the external constraint energy.

The internal energy and the image energy can be written like:

$$E_{internal} = (\alpha(s)|V_s(s)|^2 + \beta(s)|V_{ss}(s)|^2)/2 \quad (8)$$

$$E_{image} = -|\nabla I(x,y)|^2 \quad (9)$$

Where $|V_s(s)|^2$ represents the elasticity, and $|V_{ss}(s)|^2$gives the parametric curve curvature which forms the deforming snake. In the other hand, the parameter $\alpha(s)$ controls the 'tension' and $\beta(s)$ controls the 'rigidity'.

For the image energy, $\nabla I(x,y)$ denotes the gradient of the image I at (x,y).

4 Experimental results

The experiments were performed on two face databases; the first is Utrecht ECVP [8] Fig. 6, which contains 131 images of 49 men and 20 women, in both a neutral and smile position of each. It was collected at the European Conference on Visual Perception in Utrecht in 2008. The second is the FEI database Fig. 7 [7]. It is a Brazilian face database that contains a set of face images taken at the Artificial Intelligence Laboratory of FEI in São Bernardo do Campo, São Paulo, Brazil. It contains 14 images for each of 200 individuals. All images are taken in an upright frontal position with profile rotation of up to about 180 degrees.

Figure 6. Example of images from ECVP database

Figure 7. Example of images from FEI database

We use these Databases for comparison of different facial landmarks detection methods; V.J+H and V.J+A.C. We detect facial features zones using Viola and Jones method which return the rectangles that enclose the facial features, and then we use each of the Harris and snake to localize the landmarks.

Before the landmarks detection process, we have added one step, which is the detection of facial axes Fig. 8 that goes along the features in order to get the exact location of the corners. As we know, Snake detector is made to detect the contours, so adding this step will help to get the corners of each feature.

We apply the horizontal projection of the gradient image, and we find the maximum that will refer to the eyes axis. To find the nose axis, we apply the vertical projection to the gradient of the image beginning from the eye axis, and the nose axis corresponds to the area which has the minimum skin pixel around the symmetric axis. And finally to find the mouth axis, we do the same projection as eyes, and the axis is the line that has the highest gradient below the nose axis [10].

Figure 8. Features axes detection

4.1 Landmarks detection by Harris corner detector:

This section presents the results obtained from implementation of Harris corner detector on the facial features detected.

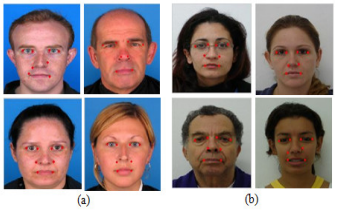


Figure 9. Results of corners detection using Harris: (a) for ECVP face database, (b) for FEI face database

The Harris corner detector detects a set of facial landmarks, among these landmarks, useless points are detected and others which are necessary are missed. Fig.9 shows some examples where the necessary landmarks are presented for the two databases.

4.2 Landmarks detection by Active contour (Snake):

The results obtained using Active contour (Snake) for facial landmarks detection are shown in Figure. 10.

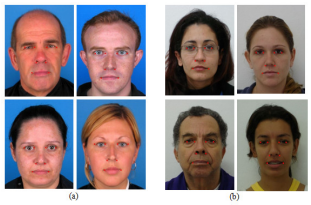


Figure 10. Results of corners detection using Active contour: (a) for ECVP face database, (b) for FEI face database

Facial landmarks detection using active contour (snake) gives very interesting results Fig.10, but the only problem is parameters initialization, which is very difficult.

In the table below, we present the success rate of facial landmarks detection for both Harris and Active contour using our two face databases Utrecht ECVP and FEI database.

Table 1. Success rate of facial features corners detection

Methods / Features	Active Contour		Harris	
	ECVP	FEI	ECVP	FEI
Eyes corners	97 %	94 %	95 %	93 %
Nose corners	89 %	90 %	81 %	88 %
Mouth corners	92 %	91%	87 %	90 %

We can conclude that the detection rate of the facial features (eyes, nose and mouth) corners by the two methods (V.J+H) and (V.J+A.C) is interesting for both methods, mainly the (V.J+A.C) that shows percentages higher than those of (V.J+H) with both ECVP and FEI databases. These results are due to the strengths of the snake detector combined with the facial axes, and the incomplete and non accurate corners detection by Harris.

5 Conclusion

In this paper, we presented two methods for facial landmarks detection using Viola and Jones method .the first combines Viola and Jones with Harris corner detector, and the second combines Viola and Jones method with Active contour (snake). Our experiments were performed on two face databases; Utrecht ECVP and FEI database, and they show that (V.J+A.C) gives very good results and had a higher landmarks detection comparing to (V.J+H).

REFERENCE

[1] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001, vol. 1, pp. I–511.

[2] Y. Freund and R. E. Schapire, “A desicion-theoretic generalization of on-line learning and an application to boosting,” in European conference on computational learning theory, 1995, pp. 23–37.

[3] D. Vukadinovic and M. Pantic, “Fully automatic facial feature point detection using Gabor feature based boosted classifiers,” in 2005 IEEE International Conference on Systems, Man and Cybernetics, 2005, vol. 2, pp. 1692–1698.

[4] C. Schmid, R. Mohr, and C. Bauckhage, “Evaluation of interest point detectors,” Int. J. Comput. Vis., vol. 37, no. 2, pp. 151–172, 2000.

[5] C. Harris and M. Stephens, “A combined corner and edge detector.,” in Alvey vision conference, 1988, vol. 15, p. 50.

[6] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” Int. J. Comput. Vis., vol. 1, no. 4, pp. 321–331, 1988.

[7] D. C. E. Thomaz, “FEI Face Database.” [Online]. Available: http://fei.edu.br/~cet/facedatabase.html. (accessed 17.03.25)

[8] P. Hancock, “Utrecht UCVP database.” [Online]. Available: http://pics.psych.stir.ac.uk/2D_face_sets.html (accessed 17.03.25)

[9] R. Lienhart and J. Maydt, “An extended set of haar-like features for rapid object detection,” in Image Processing. 2002. Proceedings. 2002 International Conference on, 2002, vol. 1, pp. I–900.

[10] F. Abdat, C. Maaoui, and A. Pruski, “Suivi du gradient pour la localisation des caractéristiques faciales dans des images statiques,” in 21° Colloque GRETSI, Troyes, FRA, 11-14 septembre 2007, 2007

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Inference-Based user's Recommendation in E-Learning Systems

Youssef Elouahby, Rachid Elouahbi
Faculty of Sciences, Moulay Ismail University, Meknès, Maroc
Elo.yusuf@gmail.com; elouahbi@yahoo.fr

ABSTRACT

This paper proposes a technique of user's recommendation for E-learning systems, which makes it possible to identify the best qualified profiles in a given field, the method is based on artificial intelligence in order to make connection between the knowledge expressed explicitly on a learner profile and a special need of another learner, not necessarily expressed on that profile, but which can be deduced through mechanism of inference.

Keywords—E-learning; recommendation of users; artificial intelligence; Inference ; Semantic Web.

1 Introduction

Face to face learning is often based on networks of mutual help between learners, the teacher intervenes only on specific cases, but in most cases, learners help their colleagues to understand concepts or solve problems, the identification of users whose can give support in classroom is easy, given the reduced number of learners, and the interactions which take place in classroom, with a simple reflection, one will know which learner is qualified on a given subject, and thus give support; on the other hand, in the E-learning systems, a learner does not have high visibility on his colleagues and their knowledge or skills, this is why we need to think of a mechanism, allowing to a struggling learner to identify someone that can help him.

To solve this problem, we use artificial intelligence techniques, to add a layer of intelligence similar to that of humans to E-learning systems, to do this, we need a semantic representation of learner's profiles.

This approach allows to learners to have a quick answer, instead of waiting for the intervention of their teacher, which can delay, it also allows to teachers to focus on more important tasks instead of intervening on every issue.

In Section 2, we present some related work to the recommendation systems of profiles, as well as some semantic web techniques and their uses in E-learning systems.

In Section 3 and 4, we present the general architecture and the implementation techniques of our profiles recommendation system.

DOI: 10.14738/tmlai.54.3198
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3198

2 State of the Art

2.1 Users Recommendation

Nowadays users recommendation in E-learning systems is a necessary technique to enable learners to understand or solve problems without necessarily going through their teachers, it allows learners to have a fast response and avoid waiting a long time before getting a response from their teachers, or even worse, sometimes a learner hesitate to ask his teacher for some reason, this deadlock situation deteriorates his motivation, and so another need of monitoring of learner's motivation appears[1], on the other hand, teachers can focus on more important tasks instead of answering each learner.

Silvan ACIAR V[2] proposes a method for user's recommendation on the basis of a Text mining technique in order to extract a user knowledge from his interaction on forums.

Fabian Abel[3] shows the importance of discussion forums in E-learning and proposes a recommendation approach to facilitate the localization of the relevant information to each learner.

2.2 Semantic Web to assist E-learning users

The new generation of the web, also called semantic web provides communication between man and machine based on a semantic language, it uses several technologies that are stacked in layers and each one exploits the capabilities of those who are below such as RDF, OWL or SPARQL.

E-learning is among the areas directly affected by the innovations of the web and should fully benefit from its advantages.

In this sense, Vibhavari R. Pandit[4] presents an approach for implementing an E-learning scenario, based on semantic web technologies and focusing on benefits of ontologies using for E-learning resources description.

Ioan Szilagyise[5] presents a semantic learning system that allows filtering internal or external educational resources (i.e. Internet) to this system. This filtering is especially based on the selection of learning objects whose annotations accredit a rapprochement with the learning profile and weaknesses of the learner.

Yi Li[6] proposes a personalized semantic recommendation system for E-learning, this system allows to semantically recommend resources for learners.

Although these works are very interesting, and have an enriching contribution to traditional E-learning systems, it remains that the semantic web is not very well exploited, especially on adaptive learning and resources recommendation.

3 Objective and Implementation Techniques

The aim of our work is to realize a user's recommendation system, who can answer one or more questions on a specific area, the novelty of this system is the way of profiles filtering and selection, which is semantic and tends more towards intelligence and human reasoning.

To do this, we must first think of a semantic modeling of the learner profile, domain and application, practically speaking building a learner profile, domain and application ontology; then we define inference rules to infer implicit knowledge of learners, and finally to exploit this knowledge, we use a query language such as SPARQL or Snap-SPARQL, this process is shown in figure 1.

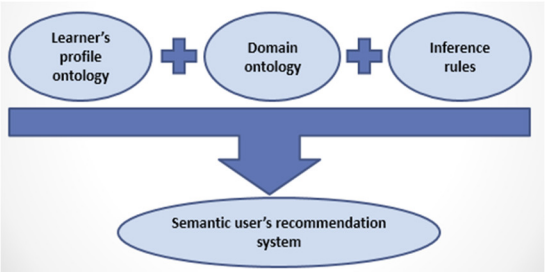


Figure 1.Semantic user's recommendation system architecture

3.1 Semantic modeling of a learner profile:

In this work, we are interested in modeling the learner's information, the purpose of the modeling is to give a complete description of all aspects relating to behavior and knowledge of the learner. At the basis of this information, we have the opportunity to do the filtering operation in order to identify the correct profile for each request, in general, learn profile modeling means collecting information about the path of this learners and his knowledge and skills. This procedure is not a goal in itself, but it is a step for performing operations such as learner's recommendation or orientation.

Adil Korchi[7] proposes a modeling approach of the learner based on an ontology. It is composed of several elements, including:

- Behavior : It is the relationship that the learner must develop with his/her environment for the success of his/her pedagogical activities. Behavior in our case concerns the motivation, autonomy, attention, responsibility, mind-opening and curiosity of the learner.
- Knowledge : This is the cognitive background of the learner. It can be general, theoretical or practical.
- Skills : Each learner admits strengths related to his/her skills. This may affect the sense of contact, creativity, speed and logic of understanding and problem solving.
- Interaction : The learner is an integral part of the interaction, otherwise he would not be a learner. It is the exchange that can have a learner with a learning environment. It may be long, medium or fast.

These are the most relevant and widely used elements in most E-learning systems, each one must play a part in our user's recommendation system, with different priorities, the first criterion that we consider, is the knowledge of a learner, being clear that the first criterion for someone to help on a given subject, is the mastery of the subject; The second criterion is skills, some issues or problems require besides knowledge, skills such as logic or sense of creativity, without taking these factors into account, and even in the presence the necessary knowledge, the resolution of these problems does not lead to satisfactory results, the behavior and interaction of a learner are low-impact elements, but they must be considered in achieving of such system, given a user who has quick interaction with the system is more likely to be present at time of request and therefore respond in a short time than a profile who has a slow interaction.

To summarize, user's recommendation is based on Knowledge, Skills, Behavior and interactions, as show in equation 1.

Users Recommendation = F (k,s,b,i) (1)

3.2 Domain ontology

The domain ontology is a crucial element for the good functioning of our recommendation system, in fact, It is based on this ontology that we can make connection between the object or concept of the request, and knowledge mastered by the recommended learner, Also this will allow us to deduce new knowledge mastered by a learner, without being already mentioned on his profile, Chaput Brigitte[8] presents an application ontology for "Applied Mathematics" module, the purpose of their modeling is to organize the notions and concepts regarding this module.

In general, the domain ontology consists of two components:

- A general ontology which consists of various general concepts in a given formation.
- A specific ontology to each course, which consists of specific notions and concepts related to this course ; in this paper, we decided to realize an ontology for Computer science course.

Computer science is a very generic and very rich area, new knowledge constantly appear, making modeling very complex task, and will always be subject to change and update according to the appearance of new concepts, However, the general organization is almost constant.

In this paper we adopt the computer science curriculum 2013 "computer science curricula 2013"[9] as the reference of our ontology, CS2013 is a report approved by the Association for Computing Machinery (ACM) and IEEE Computer Society. It is organized into a set of knowledge domains, each of which is composed of several elements and so on, there are 18 knowledge domains (Table 1), and more than 200 keywords corresponding to the thematic domains of study in Computing.

Table 1: Computer Science knowledge areas.

Knowledge Areas	Knowledge Areas
Algorithms and complexity	Networking and Communication
Architecture and Organisation	Operating Systems
Computational Science	Platform-based Development
Discrete Structures	Parallel and Distributed Computing
Graphics and visual computing	Programming language
Human-Computer Interaction	Software Development Fundamentals
Security and information assurance	Software Engineering
Information Management	Systems Fundamentals
Intelligent Systems	Social and Professional issues

We use "Protégé 5" to edit CS ontology as shown in figure 2, there are four level, and more than 250 terms, which should reflects the very fine knowledge of learners.

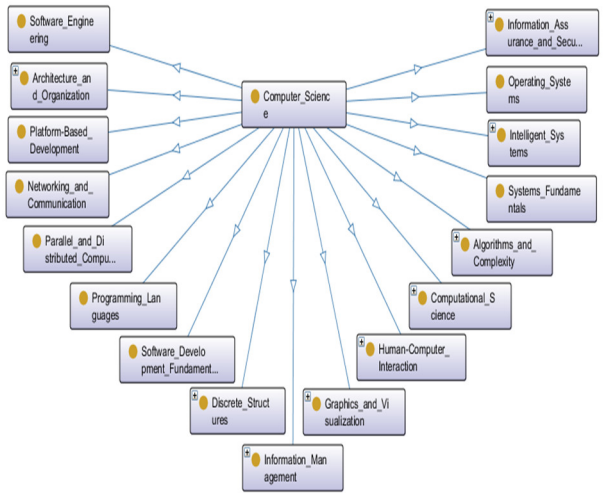


Figure 2. Computer Science ontology

3.3 Deduction of knowledge and inference engine:

Artificial intelligence allows to use a computer smartly, by facilitating communication between man and machine, and making best use of the processing power; One of the major areas of IA is the design of expert systems that are designed to meet the intelligence and performance of human experts in limited areas by exploiting a set of knowledge acquired essentially from these experts, reasoning on this knowledge requires adequate representation of such knowledge, and production rules of the form: IF condition THEN conclusion, these rules are exploited by an inference engine to produce new knowledge, this process is called inference[10].

In education field, most of notions or concepts, are composed of several concepts which are not derived from the same domain, for example we use proper mathematical concepts or sometimes physical in computer science, whence the need to design inference rules to judge some profiles able to intervene and give support about a topic, even if this particular topic is not mentioned on their profiles as a knowledge mastered, but based on the that these profiles have mastered the concepts needed to assume this concept, for example, a learner who masters the development language Java, at the same time, he has advanced concepts on network, thus our inference engine must determine that this learner is able to intervene on the subject of networks programming in Java or Socket, even in the absence of these two concepts on his profile, while a traditional user's recommendation system will not make that link.

3.4 Profiles filtering and SPARQL query:

The ontological presentation of learners profile and domain greatly facilitates filtering and localization of users mastering the desired theme, Several tools are available for this purpose, including SPARQL[11], a query language and a protocol to deal with RDF data available on a given network, it is derived from SQL, and they have many similarities, the major difference is that with SQL, we access data from a database,

whereas SPARQL gives access to all data on the web, this feature will open up the horizon to make this profile filtering technique, operating not only in a local E-learning platform, but on the global level.

4 Operating Modes

User's recommendation system operates in two modes, a simple mode, through which a query in SPARQL language browse the learners profiles and returns the users mastering the required knowledge, in the case where the system finds no Profile corresponding to the search terms, the system switches to the second mode; Intelligent mode, in which the reasoning engine is used to produce new knowledge, inference mechanism, after that a SNAP-SPARQL query is executed, this one allows to return not only the entered information but also the inferred one ; Result of the inference.

4.1 Simple query :

In the case of a simple query, the user requests assistance on a given subject, user's recommendation system sends a SPARQL query to the OMS, the result of this query is a list of users who master the subject in question, The user's recommendation system classifies returned profiles according to skills, interaction and behavior, then returns the list of sorted learners who can give support.

4.2 Intelligent query :

The simple query cannot always yield to satisfactory results, sometimes the requested theme is not written on any learner's profile, In this case the inference engine are used, and thanks to inference rules it deduces new inferred knowledge.

Unlike SPARQL, Snap-SPARQL[12] allows us to return the inferred information, which is in our case, inferred list of learners mastering the theme in question. Then the system proceeds to the classification of this list and returns it to the learner.

4.3 Multi-mode :

User's recommendation system operates in multimode, even if the two modes are asynchronous, firstly the system browse the profiles, And if this search returns no profile, inference engine are used, this transition is done by a test, which takes as argument the result of the first query.

5 Conclusion

Recommendation of learner profile who can provide assistance in case of difficulty to understand a concept or solve a problem, is one of the issues that has not been solved yet in E-learning systems, the student has no clear vision on his colleagues skills or knowledge, and even with all the information on the learners profiles, the process remains incomplete, since it often needs to add a human reasoning to find the right profile.

Our work proposes a new approach that meets this need by leveraging artificial intelligence that revolutionizes most areas, particularly E-learning.

This preliminary work has helped to design a general architecture of our profiles recommendation system, which will be completed by the conception of inference rules.

ACKNOWLEDGMENT

This work is done by the support of CNRST of MOROCCO under the Research Excellence Scholarship Program.

We would like to warmly thank the members of our research team and the scientific committee that will evaluate our paper.

REFERENCES

[1] S. Chehbi, R. Elouahbi, and F. Elkhoukhi, "Trachking of learner's motivation and involvement in moodle," Int. J. Multi-disciplinary Sci., vol. 2, no. 2, pp. 22–28, 2016.

[2] S. V Aciar, G. I. Aciar, C. A. Collazos, and C. S. González, "User Recommender based on information from forums," Proc. XV Int. Conf. Hum. Comput. Interact., pp. 3–4, 2014.

[3] F. Abel, I. I. Bittencourt, E. Costa, N. Henze, D. Krause, and J. Vassileva, "Recommendations in online discussion forums for E-learning systems," IEEE Trans. Learn. Technol., vol. 3, no. 2, pp. 165–176, 2010.

[4] V. R. Pandit, "E-Learning System Based on Semantic Web," 3rd Int. Conf. Emerg. Trends Eng. Technol., pp. 559–564, 2010.

[5] I. Szilagyi, F. Greffier, and J. Domenget, "Apprentissage personnalisé via le web sémantique," Atelier" Personnal. l'apprentissage quelles Approch. pour quels besoins?", EIAH 2011, pp. 1–8, 2011.

[6] Y. Li, L. Mei, and J. Wang, "A Personalized Recommendation System in E-Learning Environment Based on Semantic Analysis," Int. Conf. New Trends Inf. Sci. Serv. Sci. Data Min., pp. 802–807, 2012.

[7] A. Korchi, N. Elmrani Elidrissi, A. Jeghal, L. Oughdir, and F. Messaoudi, "A modeling learner approach in a computing environment for human leaning based on ontology," Int. J. Comput. Eng. Technol., vol. 6, no. 9, pp. 21–31, 2015.

[8] B. Chaput, A. Benayache, C. Barry, and M.-H. Abel, "Une expérience de construction d'ontologie d'application pour indexer les ressources d'une formation en statistique," Actes des Trente-sixièmes Journées Françaises Stat. (SFDS'04), Montpellier, Fr., 2004.

[9] ACM, Computer Science Curricula 2013. 2013.

[10] J. Haton, "L ' intelligence artificielle," 263e conférence l'Université tous les savoirs donnée, 2000.

[11] B. Quilitz and U. Leser, "Querying Distributed RDF Data Sources with SPARQL," pp. 524–538, 2008.

[12] M. Horridge and M. Musen, "Snap-SPARQL: A java framework for working with SPARQL and OWL," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9557, pp. 154–165, 2016.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Identification of Objects by Machines using RFID Technology Identify Objetcs in Internet of Things

Douae Salim, Larbi Setti, Abdellatif Elabderrahmani
Advanced Technology Laboratory, Abdelmalek Essaâdi University
Larache, Morocco
Douae.salim@gmail.com, slarbi@hotmail.com, elabderrahmani@yahoo.fr

ABSTRACT

Embedded Systems and Internet of Things know recently a revolution in terms of innovation. With connectivity and networks, embedded systems are becoming more communicative, intelligent, and autonomous. They had the ability to calculate, process information, and perception. Digital connectivity to physical objects requires, firstly, the identification of objects, in order to be able to recognize each object in a unique way and to collect the data stored, ensuring security and confidentiality. In this paper we will examine the RFID technology (Radio-frequency Identification), which enables wireless interaction over certain frequency of RFID readers with a network system, to uniquely identify, track and capture informations at varying distances without the need of human interaction. Typically, RFID readers, which emits electromagnetic waves at a certain power, in order to activate the RFID tags, if it presents in the reader's transmission range. RFID tags respond to the reader's request by emitting radio waves back with the data stored the chip. Furthermore, we will explore Identification algorithm and communication techniques between reader and tags to avoid collision, whenever a large volume of tags must be read together, or perform a transmission in the same RF field at the same time. In addition, we will describe communications protocol between reader and the information system.

Keywords-Internet of things, embedded systems, RFID, Tags, reader, radio-frequency.

1 Introduction

Identify physical objects is the first step we should perform in Internet of Things, several identification technology are available, all of them are line-of-sight, they require a close distance between reader and tag. In the other hand, RFID (Radio Frequency Identification)does not need to be very close to the objects, it is an ubiquitous automatic technology to identify and collect objects through radio frequency digital signals. The RF wave is typically electromagnetic, at least in the far field. The RF wave can also be predominantly electric or magnetic in the near field.With RFID we can detect, track, and know some specific informations about objects, it is used for many fields such as transport, health, home automation, Library, security, environment, industry.

2 RFID Physical Components:

The RFID system Composed of four main elements RFID tags, RFID readers, the air interface which define communication protocol, and edge servers. Typically, RFID readers emit radio-frequency signal that RFID

DOI: 10.14738/tmlai.54.3202
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3202

tags would detect if present in the reader's transmission range, RFID tags respond to the reader's queries by emitting radio waves back with data stored in the chip.[1] Tags can be either passive or active. A passive tag has no physical power source. It harvests energy from the reader's generated radio waves, using back scattering modulation; and consumes that energy in carrying out processing and communication tasks.An active tag has a power source and may possess certain sensing capabilities for temperature or pressure.[2]

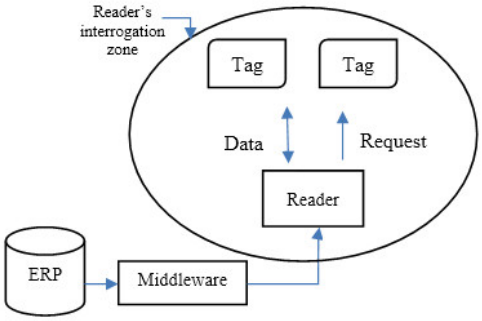


Figure. 1. Architecture of RFID System

Every tag has a unique identification code (ID). The length of ID may be different in different RFID standards; the tags in EPC class1 gen2 have 96 bit ID, but those of ISO-18000-6B (FDID, 2003) are 64 bit ID[3].It consists of a microchip programmed with information about a product and a coupling element - an antenna. As schowing in figure 1 the architecture of RFID tag is constituted by modulation / demodulation bloc, a local memory containing information about product stored in data base, and a micro-controller that represent the intelligent part of tag.

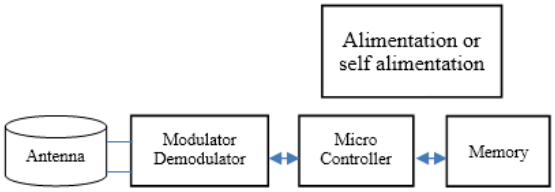


Figure. 2 . Architecture of RFID tag

Depending on the application and technology used, some interrogators not only read, but also remotely write to, the tags. For the majority of low cost tags, the power to activate the tag microchip is supplied by the reader through the tag antenna when the tag is in the interrogation zone of the reader.

Generally, the reader is constituted by (figure 2):

• An analog part regrouping:

- A local oscillator accorded on the frequency of a transmitted signal.

- A modulator/demodulator to transmit or to receive the numeric messages.

- An amplifier adapted on the antenna of emission / reception.

• A numeric part regrouping:

- A micro-controller for the management of communication protocols, collisions…

- A communication interface.

- A local memory.[4]

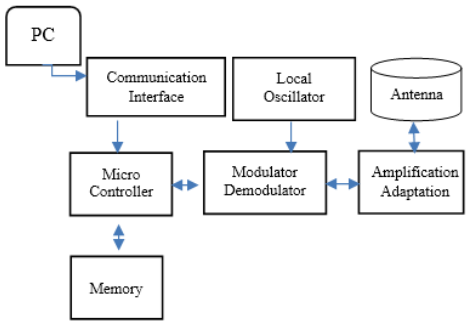


Figure . 2. Architecture of RFID integrator

In order for the chip to communicate with the reader, the tag must be open and recognized by the reader (same frequency, same speed, same modulation ...)

3 RFID Interface

The middleware RFID is the brain of the system. It manages the different readers and determines the interaction that they must have with the RFID tags. It is also the element that allows pre-processing of the captured information (filtering, routing, duplicate elimination, etc.) in order to facilitate their insertion in the databases of the information system.

4 RFID Coupling

Reader and tag communication depends on the coupling method. In RFID systems there are two type of coupling, capacitive and inductive, that allow to determine the read range and frequency of the system.The systems that use capacitive coupling use electric currents instead of the magnetic field in order to couple.

Close Coupling	Magnetic and capacitive coupling	Inductive coupling
-Range ~1cm -LF –125 kHz -Access control -tracking	-Range 1cm to 1m -HF–13.5 MHz	-Long range -Range > 1m -UHF –860MHz -Larger data storage -Backscatter coupled

5 Identification process

5.1 Communication Protocol

A communications protocol defines therules of organizing the conversation between devices in the case of RFID, between tags and a reader to ensure that information actually gets transferred correctly and efficiently. A protocol defines, air interface which contains signal modulation, collision algorithm, request and response structure, data-coding. Information security.

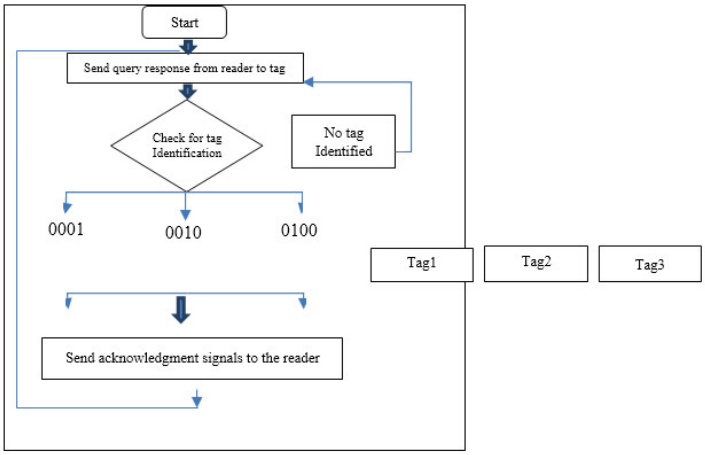


Figure. 3. Identification process

5.2 RFID anti-collision algorithms :

The anti-collision algorithm defines the way to identify multiple tags simultaneously with higher efficiency.First, we need to know the collisions categories: Tags Collisionshappen when multiple tags reply to a reader's request at the same time.Readers-to-tag Collisions happen when multiple readers readthe same tag, existing in their common range. Reader CollisionsHappen, when multiple readers are operatingat the same time, using the same frequency.

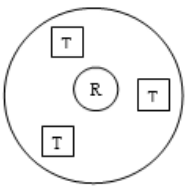


Figure. 4. Tags collision

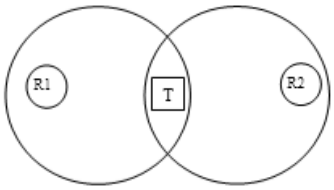


Figure. 5. Reader-to-Tag collision

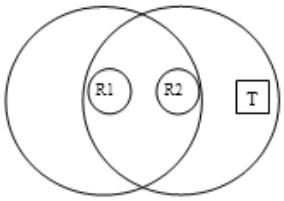


Figure.6. Readers collision

5.2.1 ALOHA Algorithms:

In RFID systems there are two most used anti-collision algorithms, ALOHA and Binary Tree algorithms based on Time Division Multiple Access (TDMA)which allows several users to share the same frequency channel by dividing the signal into different time slots.method. The reader request is divided into several slot. The reader can identify tag correctly when there is one tag responses in one slot.

ForALOHAalgorithm,therearemanyadvancedversions,forexample,SlottedALOHA,FramedSlottedALOHAandDynamicFramedSlottedALOHA.[5]

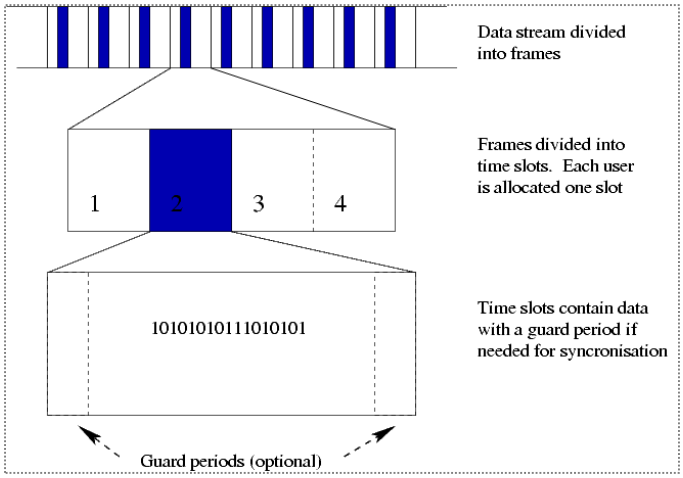


Figure. 7. TDMA frame structure

ALOHA algorithm is a simple anti-collision method based on TDMA.When the tag reaches the interrogation area of a reader, the tag will transmit the data immediately, and when more than one tag response at the same time, the collision occurs. So the most disadvantage of this algorithm is the high probability of collision.

In **Slotted ALOHA algorithm**, the time is divided into several slots, and the tag must transmit data in one slot, which it selects. So this method will decrease the probability of collision than ALOHA algorithm, but the reader and tag must communicate synchronously. When there is only one tag in one slot, reader can interrogate with tag and require the information of tag correctly. Due to the limitation of the number of slots, this algorithm used in the case that there are a few tags in the area.

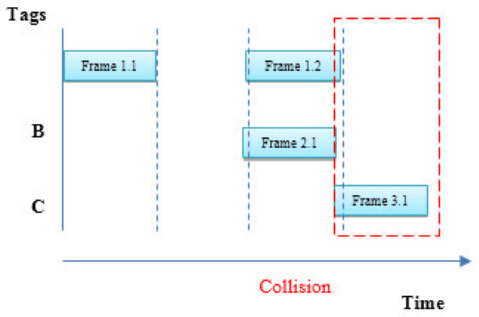


Figure. 8. Time divided into slots in slotted ALOHA

In Framed Slotted ALOHA (FSA) algorithm, one frame consists of several slots, and the tag will choose one slot in a frame to transmit data .FSA algorithm uses a fixed frame size and does not change the size during the process of tag identification. In FSA algorithm, the frame size is decided by the reader. Tags generate a random number that is used to select a slot in one frame and each tag then response in the slot it selected. Reader will identify tags with multiple frames, so it can solve the problem in slotted ALOHA algorithm.

Since the frame size of FSA algorithm is fixed, its implementation is simple, but it has a weakness that drops efficiency of tag identification. For example, in FSA algorithm, when the number of tags is small, it should choose a small number of slots in one frame, or it will cause waste of empty slots; when the number of tags is large, it should choose a large number of slots in one frame, or there will be too many collisions and it will take a long time to identify all tags.

Dynamic Framed SlottedALOHA Algorithm has more advantages than other ALOHA-type algorithms, it is proposed to improve the performance of FSA algorithm, more efficient and most widely used.In DFSA algorithm, the reader can dynamically adjustthe frame size according to the number of existing unidentified tags,

At first, the reader sends a request command to the tags and indicates the frame size. In the first read cycle,tag2 and tag3 transmit their data in slot3 simultaneously,and hence collision occurred. Tag1 and only tag1 sends itsdata in slot2, and hence the reader can successfully recognizetag1 because slot2 is singly occupied. Since there is only onetag recognizedinthe first read cycle,this implies there are still two tags that need to be recognized.Hence,the reader might determine a new contending frame size,three, for the next read cycle.

His process terminated until all tags were recognized or there is no collision occurring in one read cycle. In general, DFSA algorithm is more efficient than FSA algorithm because DFSA algorithm can dynamically adjust the contending frame size based on the current number of unidentified tags in the system.[6]

5.2.2 Binary tree algorithm:

TheBinaryTreealgorithmisbasedonatreemodel. It consists of many rounds. In each round, the reader transmits a query command and the tags respond with their IDs. The query command has a prefix. Only tags of which IDs match the prefix respond. When only one tag answers, the reader successfully recognizes the tag. When more than one tags answer, collision occurs and the reader can't recognize the tags. The reader, however, can know the existence of tags which IDs match the prefix. Then the reader tries to query with 1-bit longer prefixes in next rounds. By extending the prefixes, the reader can recognize all the tags.[5]

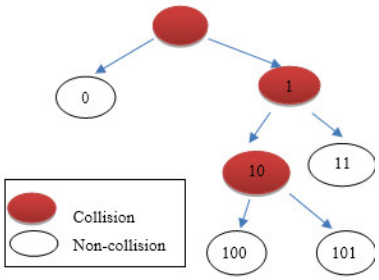


Figure. 9. Binary Tree algorithm

5.3 RFID Protocols:

There are many protocols about UHF RFIDfor exchanging information between a tag and a reader, such as EPC Class1 protocol, EPC Class1 Generation2 (Gen2), and ISO18000-6.

- EPC class1 protocol uses A bin-based Binary Tree algorithm,

-EPCClass1Gen2 isastandardizationeffort,basedon DFSA Algorithm, proposed by EPC Global. It uses techniques like frequency hopping or frequency agile systems.

-ISO18000 – 6 type A protocol uses DFSA Algorithm

-ISO18000 – 6 type B protocol uses Binary Tree algorithm

EPC Class1 Gen2

Gen2 is an air interface protocol, which defines the physical and logical requirements for an RFID system of readers and passive tags, operating in the 860 MHz - 960 MHz UHF range. It uses DFSA algorithm, and it is the most used.For Gen2 there is some operations used by reader to manage tag population, which are:Select, Inventory and Access. The reader uses “Select” to select tags population, and use “Inventory” to identify tags by transmitting a command in one of four sessions. One or more tags may reply. The reader detects a single tag reply and request the personal computer (Pc). The process “access” is used when the reader reads tag or writes to individual tag, which should be uniquely identified prior to access [12].

6 Conclusion

To remote objects in the Internet of Things,each object is uniquely identified;weproposeRFID system to identify objects.RFID is a flexible, easy-to-use technology that is perfectly suited for automatic processing, it can also provide other services such as controlling lamps, a portal ... The software integrated in the reader can be customized according to the needs of the user.

The protocols and collision algorithms are existed to correctly identify things,to ensure securityand efficiency,DFSA algorithm and Gen2 protocol are the most used to avoid collision,organize communication between items, and to standardize the RFID operation.

REFERENCES

[1] Jae-Ryong Cha and Jae-Hyun Kim, “Dynamic framed slotted ALOHA algorithms using fast tag estimation method for RFID system,” 2006, vol. 2, pp. 768–772.

[2] H. Safa, W. El-Hajj, and C. Meguerditchian, “A distributed multi-channel reader anti-collision algorithm for RFID environments,” Comput. Commun., vol. 64, pp. 44–56, Jun. 2015.

[3] M. He et al., “A fast RFID tag identification algorithm based on counter and stack,” Expert Syst. Appl., vol. 38, no. 6, pp. 6829–6838, Jun. 2011.

[4] Hamid Jabbar and Taikyeong Ted. Jeong, RFID System Integration. INTECH Open Access Publisher, 2010.

[5] I. SZENTMIKLÓSI and B. ILLÉS, “ DESIGN, OPERATION AND ANTI-COLLISION PROCEDURE OF DATA TRANSFER OF IN TEL LIGENT LOAD UNITS.”

[6] H.-W. Wang, “Efficient DFSA Algorithm in RFID Systems for the Internet of Things,” Mob. Inf. Syst., vol. 2015, pp. 1–10, 2015.

[7] A. Thomas, "RFID et nouvelles technologies de communication; enjeux économiques incontournables et problèmes d'éthique," in 6ème Conférence Internationale Conception et Production Integrées, CPI'2009, 2009.

[8] J. Myung, W. Lee, and J. Srivastava, "Adaptive binary splitting for efficient RFID tag anti-collision," IEEE Commun. Lett., vol. 10, no. 3, pp. 144–146, 2006.

[9] X. Yi, L. Wang, D. Mao, and Y. Zhan, "An Gen2 Based Security Authentication Protocol for RFID System," Phys. Procedia, vol. 24, pp. 1385–1391, 2012.

[10] http://www.enigmaticconsulting.com/Communications_articles/RFID/RFID_protocols.html

[11] https://www.etilux.be/fr/services/identification-etiquetage-marquage/tracabilite-rfid.php

[12] http://rfid4u.com/rfid-basics-resources/epc-gen2-reader-commands-and-q-parameter/

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Survey and Comparative Study on Agile Methods in Software Engineering

[1]Achrak El Mehdi, [2]Chkouri Mohamed Yassin, [1]El Kadiri Kamal Eddine
[1]LIROSA Lab, National School of Applied Sciences, TETOUAN, Morocco
[2]SIGL Research Team, National School of Applied Sciences, TETOUAN, Morocco
mehdi.achrak@gmail.com, yassin.chkouri@gmail.com, elkadiri@uae.ma

ABSTRACT

Today's business environment is very much dynamic, and organizations are constantly changing their software requirements to adjust with new environment. They also demand for fast delivery of software products as well as for accepting changing requirements. In this aspect, traditional plan-driven developments fail to meet up these requirements. Though traditional software development methodologies, such as life cycle-based structured and object oriented approaches, continue to dominate the systems development few decades and much research has done in traditional methodologies, Agile software development brings its own set of novel challenges that must be addressed to satisfy the customer through early and continuous delivery of the valuable software.

This article is located in the field of engineering Information Systems (IS) and in particular in the world of engineering process. In order, to produce information systems that the customer needs, many methods engineering were introduced. The first methods introduced are called sequential methods like Cascade method, V cycle or the Y cycle then there are iterative methods like RUP or Symphony and finally agile methods like Scrum and XP have emerged.

Keywords—Agile Methods, Development Process, Information Systems, Software Engineering.

1 Introduction

Software development is an organized process that thrives to deliver products in faster, better and cheaper ways. There have been many studies and suggestion in improving the development process. Recently, this interest has paved way to a new software development method called Agile Software Development. Agile methods strive to deliver small sets of software features to customers as quickly as possible in short iterations. As part of this paper scope most commonly used methods will examined from the angle of their applicability, strengths and weaknesses and their adoption in industry. This will lead us to find benefits, limitations and difficulties in agile software development.

The term agile stands for 'moving quickly' .Agile methodology is a lightweight methodology for software development. [1]Agile model believes that every project needs to be handled differently and the existing methods need to be tailored to best suit the project requirements. In agile development, rather than a single large process model that implemented in conventional SDLC, the development life cycle is divided into smaller parts, called "increments" or "iterations", in which each of these increments touches on each

DOI: 10.14738/tmlai.54.3203
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3203

of the conventional phases of development. Iterative approach is taken and working software build is delivered after each iteration. Each build is incremental in terms of features; the final build holds all the features required by the customer. [2]

The success of a project requires regular and frequent feedback from the client. A contract that specifies requirements, planning and project costs prior reports a vision of the IT project. The best way forward for the customer is to work closely with the development team to provide continuous feedback that ensures better control of the project. So, changes in specifications may occur very late in the project development cycle. It is a solution that truly meets customer expectations is performed, not a solution that meets the requirements of a contract established a priori. This of course requires great maturity of the customer and the service provider to establish a real relationship of trust and a good understanding of the operational reality of the project by the lawyers handling the case.

When planning, it is necessary to ensure that the schedule is flexible and adaptable to changes that may occur in the context, technologies and specifications. Indeed, it is difficult to think from the beginning that all features you would like to have and it is very probable that the customer modifies its requirements once he saw an early version of operating system.

This paper is organized as follows. Section 2 gives an overview of agile methods. In Section 3, we analyze and compare the agile methods. In Section 4, we give some recommendation to use agile methods in IS. Conclusions close the article in Section 5.

2 Agile Methodology

2.1 Agile Model

IT project management methods know as well as the technologies used a permanent challenge. The high proportion of failed projects is often fed more or less constructive feedback methods implemented. Developments in technology architectures and development tools also contribute for an important part.

The term "Agile" defines a project management approach that against the foot predictive and traditional approaches sequential cycle type V or waterfall. The very notion of "project management" is called into question in favor of "product management". In order to think more "product" that "project". After all the purpose, the project gives birth to a product.

Agile methods are mainly dedicated to IT project management methodologies. They are based on iterative and adaptive development cycles according to the customer's changing needs. They allow in particular involving all employees and customers in the development of the project.

For over a decade now, there has been an ever increasing variety of agile methods available includes a number of specific techniques and practices of software development. Agile methods are a subset of “iterative and evolutionary methods” [14,15] and are “based on iterative enhancement” [85] and “opportunistic development processes”. Most of agile development methods promote development, teamwork, collaboration and process adaptability throughout the life-cycle of the project. The major methods include eXtreme Programming [14], Scrum, Dynamic Systems Development Method, Adaptive Software Development, Crystal, and Feature-Driven Development [17].

These methods generally allow the best responding to customer expectations in a limited time (thanks in part to the involvement of the latter) while driving up employee skills. These methods are therefore a productivity gain competitive advantage as well as a client side and the provider side.

2.2 The Process of Agile Software Development

The process contains the six basic phases

- Pre project preparation
- Start
- Construction
- Production
- Selection
- Departure

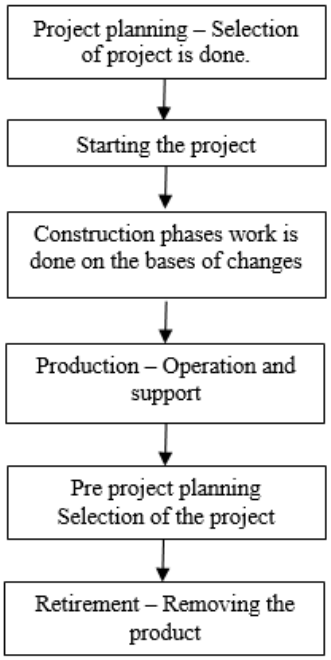


Figure. 1. Process of Agile Software Development

(1) Pre project scheduling: The basic of the project and the market features are well-defined in this phase.

(2) Start-up-phase: Requirement displaying done through active participation of stakeholders in order to identify the initial requirement basic or high-level requirements for the system. The main objective is to understand the problem.

(3) Construction: In this phase excellent working software is delivered incrementally, which meets the variable needs of the customer. Adjustment in the requirements is allowed to meet the exact needs of the customers. At the end of each progress iteration there is a fractional, working system which is shown to the customer and testing can be done on that like system integration testing.

(4) Production: It handles all operations and provides supports.

(5) Selection: The goal of the construction phase is to keep systems useful and active even after the product has been deployed to the user.

(6) Departure: It is also known as system deactivating phase.

3 Study of Agile Methods

3.1 eXtrem Programming

The main purpose is to reduce the costs of change. In traditional methods, the needs are defined and often set at the beginning of the computer project which increases costs subsequent for changes. XP works to make the project more flexible and open to change by introducing basic values, principles and practices. [4]

The client (project owner) pilot project, the team around the project delivers early a first version of the software, and deliveries of new versions then linked together at a steady pace for maximum feedback on the progress of development. The team organizes itself to achieve its objectives by encouraging maximum cooperation between its members. The team sets up automatic tests for all the features it develops, its guarantee a very high level of robustness. Developers are constantly improving the internal structure of software that changes are still easy and quick. [6]

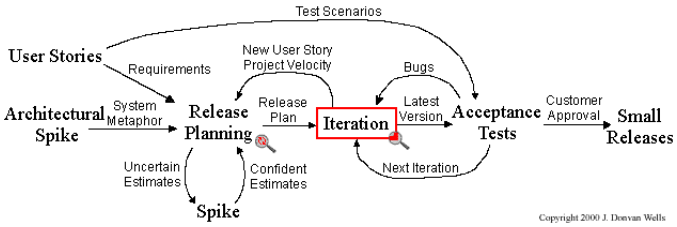


Figure 2 Overview of XP

The XP values are: communication, simplicity, feedback, respect

3.2 Scrum

Scrum is a framework for development of multifunctional teams perform iterative and incrementally products. Scrum structure development in work cycles called sprints. These iterations never last more than four weeks (two weeks most of time) and are linked one after the other without interruption. Sprints are time-limited or “Time box” they end on a specific date that the job is done or not, and are never extended. Scrum teams generally choose duration of sprint and maintained during the project, until they could further increase their productivity and then use a shorter cycle. At the beginning of each sprint, a cross-functional team (about seven people) selects items (customer requirements) in a prioritized list. The team collectively agrees on a target consisting of what expects to deliver at the end of sprint, tangibly and effectively “finished”. No new elements are added during the sprint, scrum accept change for next sprint, but the fixed term of a current sprint is made to focus on a relatively stable objective, clear and limited. [3]

Every day, the team met briefly to monitor his progress and adjust the next steps required to complete the remaining work. At the end of each sprint, a review is organized with parts during which the team shows what it has achieved. The feedback received may be taken into account on the following sprint. Scrum insists to deliver a functioning product at the end of each sprint, and really finished. In the case of software, it means an integrated, fully tested, documented for the users and potentially deployable.

The roles, artifacts and key events are show in the overview below:

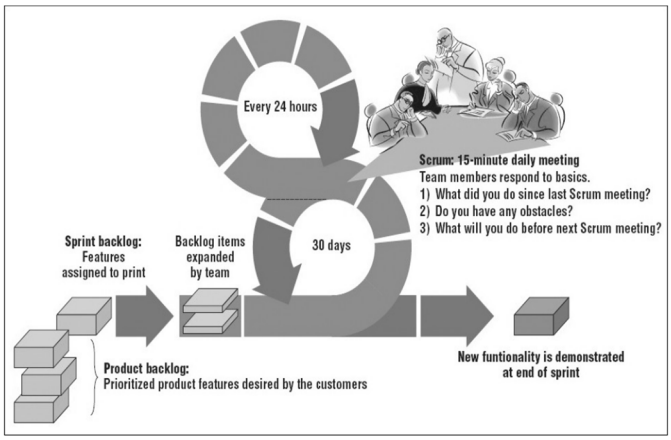


Figure 3 Scrum overview for Agile Software Development

In scrum, there are three roles: [5]

1) Product Owner: is responsible for maximizing the return on investment by identifying features of the product, transferring them in a prioritized list, selecting those to be top of the list for the next sprint, and re-prioritizing and refining continually this list.

2) Team: built product that is defined by the product owner: an application or website for example. In scrum, the team is "multi-function:" it includes all the expertise necessary to provide a version of product deliverable in each sprint.

3) Scrum Master: helps the groups learns and supply Scrum so that business value materializes. The Scrum Master is doing everything in its power to assist the team, the product owner and organization to succeed the project. The Scrum Master is not the manager of the team, or even project manager, a team leader or a representative of the team. Its role is to serve the team, its helps to remove barriers, protect the team from external interferences and facilitates the adoption by the team of modern development practices.

The values for Scrum are transparency, inspection, adaptation

3.3 Rational Unified Process (RUP)

This method can be considered the less agile methods presented here, is a mixing of practices from traditional and agile methods. The principle is browse a life cycle (inspection, elaboration, construction, transition) during iteration. Each stage in the life cycle is very precisely detailed.

It's heavy approach and the investment cost of this method the reserve for projects large and medium size detailed. [7]

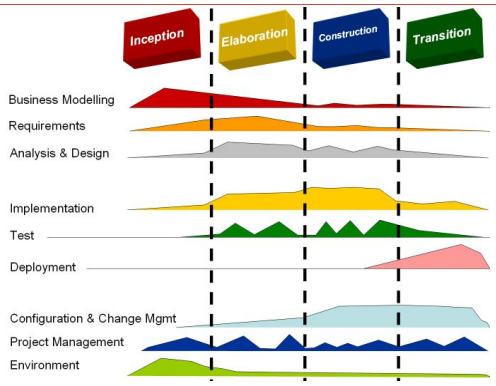


Figure 4 RUP phases

Rational Unified Process is composed of 4 phases:

Inception allows:

- Delineate the system (range), and identifying its interfaces with the external world,
- Offer or sketch a plausible system architecture,
- Identify risks regarding the feasibility of the system.

Elaboration allows:

- Develop an architecture / infrastructure to implement functionality to architectural impact,
- Identify the risks that could result in additional costs, delays, etc.
- Specify the quality requirements,
- Identify use cases,
- Make a project plan with estimated costs and resources.

Construction can:

- Complete the requirements capture,
- Complete the analysis, design, implementation and test use cases,
- Maintain architectural integrity,
- Monitor and manage risks.

Transition can:

- Prepare the deployment environment,
- Produce documentation for the 'Release' (user manuals, maintenance manuals)
- Correct errors discovered in beta.

3.4 Feature Driven Development (FDD)

Feature Driven Development (FDD), was developed by Peter Coad, Jeff Luca and Stephen Palmer. This method prioritizes risk management, based on short iterations to deliver tangible functionality to the user. This one is quickly informed of the progress, reducing the risks. According to its creators, the latter can be used in the development of critical systems.

Less known than the two previous methods, FDD focuses mainly on design and development. For this it relies on a formalization of object model using UML diagrams, cutting functions that will be developed by small teams responsible for one or two functions. It gives a very important aspect to the quality of the finished product, and support tools to monitor the progress of the project.

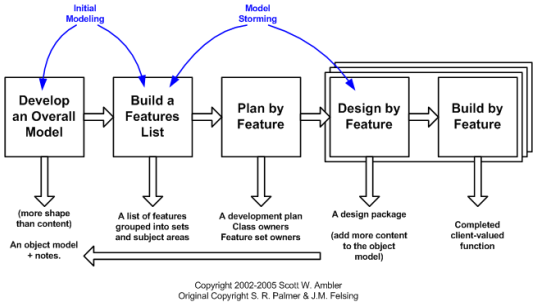


Figure 5 The FDD project lifecycle [5]

3.5 Rapid Application Development (RAD)

It is the oldest agile method and that was the first to be breaking with traditional methods. It introduced the concepts of iteration and increment. It aims to adopt the most strategic solution (in terms of time), the least risky, most reliable and least expensive. Its development cycle is simple: framing, design, construction and completion in full respect of a period of between 90 and 120 days.

The RAD cycle consists of five phases:

- Initialization determines whether the project can be followed with the RAD method,
- Expression of needs, determines the outline of the project: objectives, participants, main functions
- Design, precisely defines the processes, data,
- Construction, is the development and validation phase of the system,
- The implementation is the delivery of the newly created system.

The phases of design and construction are included in a cycle, and each is the subject of several iterations to reduce the risk of failure. [8]

This method includes 5 main steps; the last two are in the iterative cycle.

- Develop a general model.
- Build a list of features.
- Plan by features.
- Design by features.
- Build in features.

The RAD method's main objective is to produce a system in which the charge and the project period be reduced without sacrificing quality and meet the needs of users. Thus the objectives of RAD are to produce software: faster, better, cheaper.

The RAD objective is to produce a useful software when the business needs it. For this reason, the life cycle RAD is set at a maximum of 120 days. The project is contained in a short time to avoid the classic

escape ahead of the project delivery and user expectations that increase increasingly delaying project completion

3.6 Dynamic Software Development Method (DSDM)

DSDM is agile method developed in England in the mid-90s. It uses the principles already seen in other methods (user involvement, autonomy of the team, visibility and relevance of the result, iterative and incremental development, reversibility of changes, tests continuous, cooperation of actors).

The DSDM method was born of the same recognition that all agile methods, namely that development times were far too long, namely that development times were far too long, the delivered applications corresponded not always exactly the needs, the developers were very involved in the design and that no one had full view of the systems developed.

Fundamental Principles of DSDM:

- Active user involvement is imperative:
- The DSDM teams must be allowed to make decisions
- The product is made tangible as often as possible:
- An iterative and incremental development:
- Any changes during the realization is reversible
- The needs are defined at a summary level
- Testing is integrated throughout the lifecycle
- A spirit of cooperation between actors is essential.

The phases of DSDM:

- The Feasibility Study.
- The Research Business.
- The Functional Model iterative.
- Design and iterative development.
- The Implementation in the work environment.

4 Analysis of Agile methods

Agile is now widely used in web services or agencies companies. I have often heard that one of the advantages of this method was that we could have what we wanted in, but contrary to popular belief, this method will bear fruit only if it is followed to the letter. Indeed, each step is important. We will see the advantages and disadvantages and potential risks in using this method. [9]

4.1 The strenghts of Agile Methods

In this part we will present you the highlights of agile methods:

- They allow the minimization of the differences between the developed product and business needs,
- They cause a reduction in costs of study that is conducted jointly with the developments,
- They negate the possibility of tunneling to users and management, who are involved in the project,
- They are born great flexibility and responsiveness for short projects with high added value,
- The communication is better, quality is assessed continuously,

- The risks are detected early, costs are controlled.

4.2 The limits of Agile Methods

This part will allow us to know the limits which cannot exceed using agile methods among these limits:

- One of the first problems is availability. It is more difficult to obtain in small teams, including the drafting of project documentation, as everything is done progressively;
- A lack of forward planning as the project progresses custody;
- A need for consultation with other projects because the fact to all employees does not always create a team receptive to change. (Barry, January 2002)

As part of the evolution infrastructure, there may be cycles and iterations, but the user is involved or is pressed for the realization. Agility may lie in the choice of architecture (clouding, clustering, and virtualization (Boehm)) and financing options.

4.3 Gain and Downside

4.3.1 For the Customer:

Putting the customer at the center of the steps and people involved in it many advantages:

- In terms of satisfaction, because the customer can adapt, even in development, the software to its business needs by giving her feedback through regular access to different functional versions made. The goal of agile methods is indeed the real satisfaction of need rather
- In terms of competitiveness: direct adaptation needs of the software definitely provide great value to it. Management priorities in cost estimates also provide greater efficiency in economic terms for the client.

Agile methods fall short cycle as a project but resulting length that is important, so customers who have little time or interest in the proposed development of their software will provide a final product much lower in terms of quality. [10]

4.3.2 For the developer:

One of the principles of agile methods is self-organization of team development; developers enjoy greater creative freedom and greater responsibility in the project. In addition, the prominence of trade and the collective spirit promises better quality of working life. For their managers, the advantage is greater visibility on project progress.

Finally, management will appreciate the focus on return on investment and the sustainability of the program throughout its lifetime.

However, by focusing on direct dialogue, agile methods leave little room to documentation, which can become a big problem if development team when changes or software maintenance is handled by another team than which has been responsible for the development. [11]

4.4 Comparaison between Agile Methods

After listing the agile methods now we go to make compare between these methods. We will start by seeing the percentage of use, development phases, advantages, disadvantages, size of the project, cost of development, rapid of use and quality for each method.

Percentage of use:

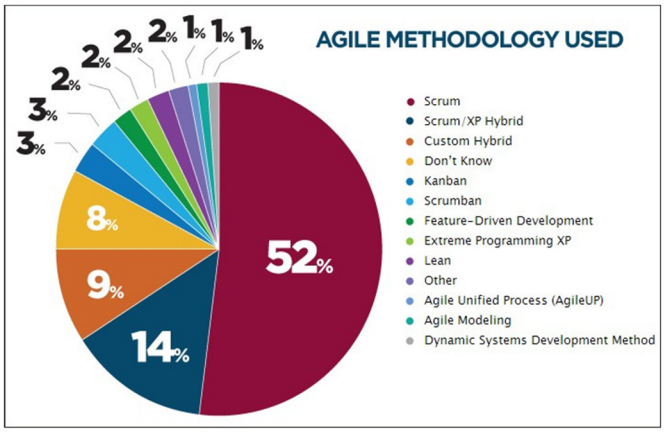


Figure 6 Li State of Agile Survey Results 2011 by Version One In [13]

Development phases

- Scrum : Vision / Product Backlog / Sprint Backlog / Product Increment
- XP : Planning / Writing tests / Development / Test / Integration
- RUP : Inception / Development / Construction / Transition
- FDD : Development / Build / Plan / Designate
- RAD / DSDM : Initialization / Expression of needs / Design / Construction / Implementation

Advantages :

- Scrum : fully developed and tested for short iterations, simplicity of processes, clearly defined rules, increased productivity, personal organization, each team has its share of responsibility and improved communication.
- XP : reduce costs of change, permanent code reviews, development driven by testing, design during the project and systematically privilege the simplest method.
- RUP : risks are mitigated from the start of the project, change is better managed, improved portability through the use of UML and verify the constant quality.
- FDD : risk management, frequent deliveries and code Inspection.
- RAD : rapid development, low development cost and a high-quality application.
- DSDM : quick implementation of priority solutions, adequacy of the system to the need, respect of deadlines and costs and integrated tests

Disadvantages:

- Scrum : little written documentation and violation of liability.
- XP : small and medium-sized projects only, requires strong customer involvement and not enough documentation.
- RUP : heavy management to set up and not suitable for a small project.
- FDD : number of senior developers available and code inspection.
- RAD : applications are very long to develop.
- DSDM : difficult maintenance and bad understanding with users.

Size of the project:

- Scrum : many processes implicitly define.
- XP : 12 people in the team.
- RUP : importance of well-defined processes.
- FDD : 20 people in the team.
- RAD : presence of a coordination and reporting group.
- DSDM : extensible due to roles and well-defined responsibilities.

Cost of development :

- Scrum : $633 043
- XP : $630 860
- RUP : $756 157

Rapid :

- The deadlines, the costs.

Quality :

- The potential for anomalies
- Efficiency in repairing anomalies (DE)
- Anomalies present during delivery
- Major anomalies

After the comparison between the agile methods, we can say that the two ideal methods that can be used : XP and Scrum. So we will do a detailed comparison between these two methods. [12]

XP Method	Scrum Method
Duration of the iteration (1 to 2 weeks)	Duration of the iteration (2 to 4 weeks)
The changes during the iteration features are not accepted.	Possibilities to changes opportunities scenarios or features to implement in the iteration.
Different roles are assigned to team members 'XP' (programmer, tester, coach, manager, ...)	The team members are multidisciplinary and only three roles are defined (scrum master, product owner and the team)
Management practices, engineering practices and tools for collaboration and the support of management	Management practices and tools for collaboration and the support of management

Scrum following an iterative process for obtaining a product of close customer needs by taking into account the evolution of the latter and thus maximize the value of the delivered product.

If during a Sprint, a problem occurs, the responsibility lies not with one person but is shared between the Product Owner, the Scrum Master and Scrum team.

4.5 Recommendation for use agile methods in SI

To have a well-structured information system and especially without any problem agile methods are used according to our needs and the project.

For this we made a vision and we follow three steps:

- First step: formalize the product vision: objectives, target users
- Step two: establish a list of functional and non-functional requirements
- Third step: estimating each requirement by the development team

Requirements will be converted into usable features as scheduling. [9]

So you have to back up the functional requirements of the highest value at the top of the list.

One can also use a planning poker: it is a technique to rapidly reach an optimal and objective estimation. Before or during the estimates, the product owner may be solicited to answer the questions of the development team. To conclude, we can say that Scrum offers several advantages with the possibility to apply on a large project panel. It is for everyone to have their own idea and decide if the Scrum methodology is adapted or not to this or that project because it's important to keep in mind that each team, each project and context are different.

5 Conclusion

Agile approaches are increases the flexibility, agility and to be more adjusted to the environment where software development projects are present and working today. The idea behind Agile methodology would likely to break large projects into smaller projects which would become more flexible. Also it's proven that agile provides means for software development with minimum risk where project requirements are well defined, development team is self-dependent. It also ensures that every team member understands, evaluates and actively participate in development which ensures no single point of dependency and failure. It also facilitates more flexibility to adopt changes in requirement and control development life cycle through micro breakdown of requirements and planning of same.

REFERENCES

[1] K. Beck, Extreme Programming explained: Embrace change. Reading, Mass., Addison-Wesley, Nov16, 2004

[2] Jim Highsmith "Agile Project Management:creating Innovative product" (2nd Edition)

[3] Charlotte, H. . Méthodes, modèles et outils pour la méta-modélisation des processus d'Ingénierie de Systèmes d'information. France, 2009

[4] Chausson, C. (2008). Méthodes agiles: le renouveau des relations client.

[5] Deemer, P. (2013). The Scrum Primer: Guide Léger de la Théorie et de la Pratique de Scrum - Version 2.0.

[6] Jean-Louis, B. (2011). Extreme Programming - Agile Methods . Tour d'horison / Business Interactif.

[7] Koch, A. S. (2004). Agile Software Development Evaluating the Methods for Your Organization. ARTECH House.

[8] Messager, V., & Tabaka, J. (2007). Gestion de projet vers les méthodes agiles . Groupe Eyrolles,.

[9] Paasivaara, M., & Lassenius, C. (s.d.). Could Global Software Development Benefit from Agile Methods? IEEE International Conference on Global Software Engineering , 5.

[10] Palmer, S. R. (s.d.). A Practical Guide to Feature-Driven Development 1st Edition.

[11] Rahmania, a. e. (2012). Project Lifecycle, Project Management, tagged agile methodology, extreme programming, rational unified process, RUP, SDLC.

[12] STEPHANE, L. Transition vers l'agilité à l'échelle d'une organisation. Valtech, 2012

[13] http://www.versionone.com/Agile101/Agile-Soft ware-Development-Benefits/

[14] C. Larman, Agile and Iterative Development: A Manager's Guide. Boston: Addison Wesley, 2004.

[15] C. Larman and V. Basili, “A History of Iterative and Incremental Development,” IEEE Computer, vol. 36, no. 6, pp. 47-56, June 2003.

[16] V. R. Basili and A. J. Turner, “Iterative Enhancement: A Practical Technique for Software Development,” IEEE Transactions on Software Engineering, vol. 1, no. 4, pp. 266

[17] A Practical Guide to Feature Driven Development. 2002 S Palmer, M Felsing - Prentice Hall

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Counterterrorism: Privately Clustering A Radical Social Network Data

Jamal Boujmil, N. Tagmouti, N. Raissouni
RS&GIS Lab. Dept. Telecommunications
The National School for Applied Sciences of Tetuan
Tetuan, Morocco
jamal.boujmil@gmail.com, nohatg@gmail.com, nraissouni@uae.ma

ABSTRACT

The tradeoff between the needed or essential gathering and analysis of personal data and the privacy rights of individuals is now an important requirement under any counterterrorism program. The most famous and controversial recent example is the revelation that US intelligence agencies systemically engage in "bulk collection" of civilian "metadata" detailing telephonic and other types of communication and activities, with the alleged purpose of monitoring and thwarting terrorist activity. Differential privacy provides one of the strongest privacy guarantees up to now. In this paper, we present a new provably privacy-preserving algorithm able to identify and take action upon members of the targeted subpopulation. Meanwhile, avoiding compromising the privacy of the patriot subpopulation. It is a new algorithm for search methods which use a new combination of nodes social similarity and differential privacy.

Keywords: differentail privacy; social similarity; privacy preserving

1 Introduction

The conflict between the useful or essential collection andanalysis of data and the privacy rights as a fundamental human right of the citizens is at an historical issue. The controversial recent example is Snowden revelations that US intelligenceagencies systemically engage in mass collection of civilianmetadata. This withthe purpose of monitoring andtracking terrorist activities worldwide.

Existing models for data privacy guarantee only an"all or nothing" flavor: privacy rights are either provided toevery member of a nation, or else it is deemed to be afailure[1]. However, those models are valid only if all the populationmembers have the same privacy right or demand. Facing the cruel terrorist group's activities including "Daesh"and similar groups, privacy rights of related members should be reconsidered. It leads to a serious questions about the balance between protectingthe rights of ordinary citizens, while opting for all necessary meansto avoid terrorism. It means that we have to make a trade-offs between the protection of the civilian privacy right and scarifying it in service of such a societal priority. A recent US National Academies studyconcluded that there are not (yet) technologicalalternatives to bulk collection and analysis of civilian metadata,in the sense that such data are essential in current counterterrorismpractices[1].

DOI: 10.14738/tmlai.54.3204
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3204

Our model tends to explicitly acknowledgesuch trade-offs. Thus we will divide a population to two groups: “Terrorist: T” and “Patriot: P”. This last group is a subpopulation to protect and for which the privacy is guaranteed by law, they are to be contrasted with thefirst subpopulation group, which does not sharethose privacy assurances. The population member status whether he is a “T” or a “P” member is unknown, however it can be discovered basing on costly measures including surveillance or intelligence data exchange.

The model balance’s objective is to give a tool to intelligence agencies to identify and take necessary actions, in a large population modeled in a social, and take appropriate actions on the “T” subpopulation. Meanwhile, privacy assurances for “P” subgroup is guaranteed.

A common “contact chaining” graph search method can be used, since starting from known terrorist “seed” vertices in the social network, neighboring vertices are investigated, in order to identify the “T” subnetwork which is the immediate neighbor of the seed. The main concern of such kind of research methods is that it has access to the vertices and related edges information. This will compromise the privacy of the “P” subgroup.

Our objective is to provide a deliberately “noisy” and privacy preserving version of the search method mentioned above, which will provide the intelligence agency a list of the eventual terrorist in a social network for whom subsequent actions are needed. Meanwhile, privacy rights for the “P” will be guaranteed.

Following are the basis of our model:

i. Social network graph data comprising certain number of private data of each involved individual and which they desire to protect. We assume that the intelligence agency has direct access to this network data, and would like to discover and act upon “T” individuals.
ii. ii) A permanent “tag bit”, which defines the membership status of each individual to “T” or “P” subpopulation is used. It can be checked by the third party agency at certain cost, so, the discovery budget should be considered in order to be minimized by the model. The attribution process of the “tag bit” to each individual is unobservable and we suppose that this classification is done basing on certain investigations conducted by related competent agencies or authorities.
iii. iii) Differential data privacy mathematical model[3] which provide rigorous guarantees of privacy for network’s data for individuals for whom the data privacy is required, meanwhile, it allows the discovery for the “T” subpopulation.

It is crucial to define the “privacy” type we want to protect since that this word comprises many definitions and concepts as per Solove’s taxonomy of privacy[4]. We are concerned with the informational privacy which is based on the quantification of information we can deduct from the output of an analysis related to a protected individual.

2 Key Contribution

We make three key contributions in this paper. First, we introduce a privacy preserving class of graph search algorithm designed to find and identify “T” individuals. Second, the algorithm is based on a new network node similarity method introduced by Yong Li [5] to measure the social similarity between all network nodes.Third, since the social similarity is a kind of private date which concerns the compared nodes, thus our randomized algorithm add noise to the output of this similarity function and construct a

noisy similarity matrix. A contact chaining process will drive the search inside the similarity matrix to define the “T” subpopulation.

This framework is the first works to our knowledge to output an explicit list of protected and targeted subpopulations withqualitatively differing privacy rights basing on social similarity techniques.

3 Related Work

The notion of differential privacy and its usage in the community detection or clustering was developed in a series of papers.In [6], authors find that k-anonymization, when done "safely", and when preceded with a random sampling step, satisfies □-differential privacy with reasonable parameters. Another approach was proposed in [7], to upgrade the one mentioned in [8] which does not satisfy differential privacy. They choose Louvain method as the back-end community detection for input perturbation schemes and propose the method LouvainDP which runs Louvain algorithm on a noisy super-graph. For algorithm perturbation, they design ModDivisive using exponential mechanism with the modularity as the score. Another differential privacy community detection algorithm was presented in [9], where it has implications to private data exploration, clustering, and removal of outliers. Furthermore, they use it to significantly relax the requirements of the sample and aggregate technique, which allows compiling of "off the shelf" (non-private) analyses into analyses that preserve differential privacy. Ahmed et al.[10] propose a privacy preserving mechanism for publishing and clustering social network graph data, which satisfies differential privacy guarantees by utilizing a combination of theory of random matrix and that of differential privacy. Meanwhile, Day et al.[11] investigated the problem of publishing the degree distribution of a graph under node-DP by exploring the projection approach to reduce the sensitivity. They proposed two approaches based on aggregation and cumulative histogram to publish the degree distribution.

Another similar approach to our work is the one presented in [1], where the goal was the development of algorithms that can effectively identify and take action upon members of the targeted subpopulation in a way that minimally compromises the privacy of the protected, while simultaneously limiting the expense of distinguishing members of the two groups via costly mechanisms such as surveillance, background checks, or medical testing

One of the main characteristics of our work is the usage of the social similarity as presented in [5] to output the list of the “T” subpopulation. All the published works up to now uses different network similarity techniques for social recommendation[12] or similar purposes. No one of them use it for a differential privacy community detection purpose. For reference we present here after the main works which used the social network similarity for a community detection purpose. Authors of [13]presented a novel algorithm for community detection that combines network structure with processes that support creation and/or evolution of communities. It identifies leaders, and communities that form around those leaders. It naturally supports overlapping communities by associating each node with a membership vector that describes node's involvement in each community. This way, we can identify nodes that are good followers to their leader, and also nodes with no clear community involvement that serve as a proxy between several communities and are equally as important.

Another measure of similarities between vertices based on random walks was proposed [14] and which has several important advantages: it captures well the community structure in a network, it can be

computed efficiently, and it can be used in an agglomerative algorithm called Walktrap to compute efficiently the community structure of a network. The most important work is the one presented in [5] and which we adopt for our approach. They utilize the concept of information and information loss to measure the node similarity. The whole model is based on this idea that if two nodes are more similar than the others, then the information loss of seeing them as the same is less. The present new method has low algorithm complexity so that it can save much time and energy to deal with the large scale real-world network.

In [15], authors have studied the nearest neighbor search problem in complex network via the development of a suitable notion of nearness.

4 Differential Privacy

Differential privacy has different interpretations[3], it is a formal framework to quantifyto what extent individual privacy in a statistical databaseis preserved while releasing useful aggregate information about the database[16]. It assures privacy guaranteesby requiring the indistinguishability of whether or not anindividual data is in the database following the releaseof information after different query by an analyst. Simultaneously for every individual v's data, and in the same time for any analysis algorithm or query Q that theymight be concerned about, Q is almost no more likely tooccur given that this single individual v's data is used in the computation,compared with if it were replaced by any arbitrarily differententry. It means in a practical way that what an observer or analyst learn about an individual "Amine" (e.g., that "Amine" is in contact with another person Baghdadi, or a group of persons, such as members of "Daesh"radical group) is almost independent of Amine'sconnections, so long as Amine is not herself a member of the "T"population. However, it does not prevent himfrom learning that Amine exists at all. This models a settingin which (for example) a national government has access to anindex of all of its citizens (through birth and immigration records),but nevertheless would like to protect information about theirinteractions with each other[1].

Differential privacy definition is introduced here under. It is oriented for social networks graph dataG where individuals are represented as vertices and the social interaction between them is represented as edges (weighted or unweighted). The information which comprises an edge is a private information to protect.

4.1 Definition 1[3]

An algorithm A is □-differentially private (□-DP) if for all pairs of neighbors graphs G, G' and for every output in the range of A:

$$Pr[A(G) \in S] \leq e^{□} Pr[A(G') \in S] \quad (1)$$

Network graphs G and G' are neighboring graphs if one can be obtained from other by:

- Adding or deleting one edge from one of them or by arbitrary rewiring of the edges incident to a single vertex: *Edge differential privacy*,
- Adding or deleting a node and its adjacent edges: *Node differential privacy*

The smaller the risk multiplier $e^{\varepsilon}$is, the more guarantees to individual data are assured. Its value is directly related to the value of the privacy parameter or budgetε. Hsu et al.[17] deeply study the "economic" alternatives for choosing ε**.**

4.2 Laplacian mechanism

Since its introduction in [18], the Laplacian mechanism has become the standard technique to let a mechanism or algorithm to achievedifferential privacy and has beenused as the basic building block in a number of works on differential privacy analysis in other more complex problem settings[19]. It is based on the concept of *globalsensitivity* of a function fwhich is defined as:

$\Delta f = \max_{G,G'} // f(G) - f(G')||_1$where the maximum is taken over all pairs of neighboring G, G'. Given a function f and a privacy budget ε, the noise is drawn from a Laplace distribution:

$Lap(\lambda) : p(x| \lambda)=(1/2l)e^{-|x|/\lambda}$where $\lambda= \Delta f /\varepsilon$.

A. Theorem 1: Laplace mechanism[18]
For any function f: G →IR^d, the algorithm A:

$$A (G) = f(G) + Lap_1(\Delta f /\varepsilon),\ldots, Lap_d (\Delta f /\varepsilon) \quad (2)$$

Satisfies -differential privacy, where $Lap_i(\Delta f/\varepsilon)$ are Laplace variables with scale parameter $\Delta f /\varepsilon$

5 Framework

In this section, we will propose our detail differential approach to output the list of the "T" subpopulation in a social network**.**

Let's considerG as a binary graph representing the connectivity of the social network, and A be the adjacency matrixrepresenting the graph, where $a_{i,j}$is the weight between two different nodes i and j. By assumingthat the graph is undirected, A will be a symmetric matrix,i.e. $a_{i,j}= a_{j,i}$for any i and j.

The first step of our framework is the algorithm 1 called DPSim: Differential privacy social similarity output. It is mainly about calculating the social similarity of the graph G basing on the adjacency matrix A. The key element at this stage is the usage of the new social similarity measure introduced by Yong Li [5] where the basic idea is that if two nodes are more similar than the others, then the information loss of seeing them as the same is less than that of others. We will adopt the proposed algorithm in [5] which output the node similarity measure between two nodes i and j based on the information loss proposed in the same work. The larger value of the algorithm output means higher similarity between the corresponding two nodes. This measure is a critical data to protect against any inference, thus we will opt for a perturbation to each entry of the similarity matrixS. To do that, we will generate a Gaussian random matrix $Q \in IR^{n*n}$and where each entry of Q is sampled independently from another Gaussian distribution $N(0, \sigma^2)$. The value of σis defined in [10] and finally the perturbed similarity matrix Š = S + Q.

The following table describes DPSim algorithm:

Algorithm1 : **DPSim (G, A, σ^2)**

Input:

1) **Social network graph G**
2) **Symmetric adjacency matrix** $A \in IR^{n*n}$
3) **Variance for random noise** σ^2

Output: Randomly perturbed similarity matrix Š

(1) Similarity matrix S = Algorithm (G, A) [5]

(2) Compute a random perturbation matrix Q, with $Q_{i,j} \sim N(0, \sigma^2)$

(3) Compute the randomly perturbated similarity matrix $Š = S + Q$

The second step of the framework is the algorithm 2 called DPSearch. The main idea of the algorithm is that the node which has the high social similarity score with other neighbors nodes in the same network, have the high probability to be in the same class or group of subpopulation with them. It means, if it is a "T" or "P" individual, related similar neighbors with high probability can be in the same subpopulation group. Thus, at this stage, we will use the perturbed similarity matrix Što select at each iteration the seed node with the high social similarity score to query its tag bit and discover to which subpopulation group it belongs. If it is a "T" member, contact chain process consisting of tag bit discovery is initiated for its neighbors.

The following table describes DPSearch algorithm:

Algorithm2 : **DPSearch (Š)**

Input: The randomly perturbated similarity matrix Š

Output: List of "T" subpopulation

Initialization: I = Ø , T = Ø , Count = n

(1) Calculate $Š_i = \sum_{j=1}^{n} Šij$**, for each i= 1,…n**

(2) $V = \{ Š_i / i=1,\ldots, n\}$

(3) While Count ≠ 0

Choose $S_{max} = \sup V = S_{i\,max}$

Count = Count – 1

$V = V \setminus S_{i\,max}$

Query TB(n_i) and I = I $\cup$ **{n_i}**

If TB(n_i) = 1 Then T = T $\cup$ **{n_i}**

Loop

For j = 1,…, n and j ≠ i

If $Š_{ij} \neq 0$ **and** $n_j \notin I$

Then:

Query TB(n_j) and I = I $\cup$ **{n_j}**

If TB(n_j) = 1 Then T = T $\cup$ **{n_j}**

End

End

Algorithm 1 as demonstrated by Kenthapadi et al. in [20]satisfies ε-Differential privacy basing on the following assumptions:$\delta<1/2$, $n\geq 2$ and $\sigma\geq(1/\varepsilon)[10(\varepsilon+\ln(1/2\delta))\ln(n/\delta)]^{1/2}$.

6 Experiments and Results

In this section we will introduce our planned approach to demonstrate the effectiveness and performance of our differential privacy searchalgorithm against its competitors. Up to now we are in the algorithm implementation phase, thus related experiments are not yet initiated. It will be a subject of another research paper.

6.1 Dataset

In our experiments we will use four different social network graphs. The first one is a real "snapshot" of twitter social network by DMI-TCAT tool [2]basing on specific key words or hashtags which we set up. In our case, we will capture all twitter accounts which are involved in a radical propaganda around following key words and hashtags for a limited time period: "Raqa", "Daesh", "Halab", "Jabhat Annossra", "Akhbar Asham". Those words or hashtags are chosen basing on specific events related to certain radical groups including Daesh. The tool output the social interaction between two accounts including: number of retweets, number of mentioning, and number of shared / like web pages. The social network will be modeled as following:

- Each account will be considered as a node
- An edge between two nodes will be established if they share one of the following events: retweet, mention, same page like. It is important to mention that each edge will be weighted and the weight between two nodes is the sumup of the retweets, mentioning and shared / like pages [21]

The second social network graphs group comprises three networks from Facebook, Live Journal and Pokec. For Facebook data set, we will consider the one collected by Wilson et al. from Facebook [22]. For both Live Journal and Pokec, they are publically available at SNAP graph library [23], [24]. Those social networks are chosen for the fact that theyshould be large enough to truly represent real online social structure. A small network not only under-represents the social structure, but also produces biased results. Also, the number of edges should be large to reveal the interesting structure of the network. For all three benchmark datasets, the ratio of the number of edges to the number of nodes is between 7 and 20 [10] and for the first one it is about 15.

To fulfill the main assumption of our approach which requires a predefined "Tag bit" to identify an individual membership to "T" or "P" subpopulation, and since the networks we will use do not have this identifiers, so we will opt for Pre-processing phase of those networks. It will help us to generate "Tag bit" for "T" and "P" subpopulation synthetically. We will use the "Infection" process adopted by the author in [25]

7 Conclusion

In this research, we focused on algorithm implementation phase proposing a new differential privacy approach basing on nodes social similarity to output and identify members of a target group "T". Results are highly promising providing strong guarantees about the privacy of the patriot subpopulation "P".

REFERENCES

[1] Kanski, J.J., *Clinical ophthalmology*. 6th edition ed2007, London: Elsevier Health Sciences (United Kingdom). P.952.

[2] Liang, Z., et al., *The detection and quantification of retinopathy using digital angiograms.* Medical Imaging, IEEE Transactions on, 1994. 13(4): p. 619-626.

[3] M. Kearns, A. Roth, Z. S. Wu, and G. Yaroslavtsev, "Private algorithms for the protected in social network search," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 4, pp. 913–918, 2016.

[4] S. Kok and R. Rogers, "Rethinking migration in the digital age: Transglocalization and the Somali diaspora," *Glob. Networks*, no. July, 2016.

[5] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 2013, pp. 211–407, 2014.

[6] D. J. Solove, "Conceptualizing privacy," *Calif. Law Rev.*, vol. 90, no. 4, pp. 1087–1155, 2002.

[7] Y. Li, P. Luo, and C. Wu, "A new network node similarity measure method and its applications," no. 71271070, pp. 1–12, 2014.

[8] N. Li, W. Qardaji, and D. Su, "On Sampling, Anonymization, and Differential Privacy: Or, k-Anonymization Meets Differential Privacy," 2011.

[9] H. H. Nguyen, A. Imine, and M. Rusinowitch, "Detecting Communities under Differential Privacy," 2016.

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, vol. 10008, no. 10, p. 6, 2008.

[11] K. Nissim, U. Stemmer, and S. Vadhan, "Locating a Small Cluster Privately," *Proc. 35th ACM SIGMOD-SIGACT-SIGAI Symp. Princ. Database Syst.*, pp. 413–427, 2016.

[12] F. Ahmed, R. Jin, and A. X. Liu, "A Random Matrix Approach to Differential Privacy and Structure Preserved Social Network Graph Publishing," 2013.

[13] W. Day, N. Li, and M. Lyu, "Publishing Graph Degree Distribution with Node Differential Privacy," *Proc. 2016 Int. Conf. Manag. Data*, pp. 123–138, 2016.

[14] Z. Jorgensen and T. Yu, "A Privacy-Preserving Framework for Personalized , Social Recommendations," *Proc. 17th Int. Conf. Extending Database Technol.*, pp. 571–582, 2014.

[15] A. Stanoev, D. Smilkov, and L. Kocarev, "Identifying communities by influence dynamics in social networks," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 84, no. 4, 2011.

[16] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *Comput. Inf. Sci.*, vol. 10, pp. 284–293, 2005.

[17] C. Detection, "Nearest Neighbor search in Complex Network for Community Detection," pp. 1–13, 2015.

[18] S. Oh and P. Viswanath, "The Composition Theorem for Differential Privacy," *Int. Conf. Mach. Learn.*, vol. 37, p. 26, 2015.

[19] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, and B. C. Pierce, "Differential Privacy: An Economic Method for Choosing Epsilon," *IEEE Comput. Secur. Found. Symp.*, pp. 1–29, 2014.

[20] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," *Lect. notes Comput. Sci.*, no. 3876, pp. 265–284, 2006.

[21] Q. Geng and P. Viswanath, "The Optimal Mechanism in epsilon,delta-Differential Privacy," p. 16, 2013.

[22] K. Kenthapadi and A. Korolova, "Privacy via the Johnson-Lindenstrauss Transform," *arXiv Prepr. arXiv …*, no. 1, pp. 1–24, 2012.

[23] F. Ahmed and M. Abulaish, "A Generic Statistical Approach for Spam Detection in Online Social Networks," vol. 36, pp. 1120–1129, 2013.

[24] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "User Interactions in Social Networks and their Implications," *Eurosys'09 Proc. Fourth Eurosys Conf.*, pp. 205–218, 2009.

[25] L. Takac and M. Zabovsky, "Data Analysis in Public Social Networks," *Int. Sci. Conf. Int. Work.*, no. May, pp. 1–6, 2012.

[26] J. Yang and J. Leskovec, "Defining and Evaluating Network Communities Based on Ground-Truth," *ACM SIGKDD Work. Min. Data Semant.*, pp. 745–754, 2012.

[27] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," *Kdd*, p. 137, 2003.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Collective Behavior Bees for Solving HW/SW Partitioning and Scheduling Problems in RSoC

Yahyaoui Khadidja, Bouchoicha Mohammed
Department of math, Faculty of Exact Sciences, University of Mascara, 29000 Algeria
yahyaouiinfo@gmail.com; Bouchoicham_info@yahoo.fr

ABSTRACT

In the codesign domain, many hardware and software techniques must be developed to satisfy specific constraints in terms of computation time, area, performance, power consumption, etc. This paper introduces an automatic approach To perform HW/SW partitioning and scheduling such that the global application execution time is minimized and the majority area of FPGA (Field programmable gate array) used in RSoC (Reconfigurable System on Chip) is exploited. The used algorithm is inspired by the collective behavior of social insects such as bees. : Honey Bees Mating Optimization (HBMO). Comparing the proposed method with Genetic algorithm, the simulation results show that the proposed algorithm has better convergence performance.

Keywords—RSoC, HW/SW partitioning; scheduling; Honey Bee Mating Optimization; Genetic algorithm.

1 Introduction

Using a Reonfigurable System on Chip (RSoC) is increasingly common in embedded applications. RSoC is a circuit comprising multiple functions such as one or more processors, one or more reconfigurable devices like FPGA (of Field Programmable Gate Arrays), a signal processor DSP (Digital Signal Processor), various peripherals and memory or analog parts. These circuits are increasingly used because of their small size and reduced costs compared to the use of various circuits for performing the same function. However, the use of Field Programmable Gate Arrays (FPGAs) provides specific hardware technology, which can also be reprogrammable thus providing a reconfigurable embedded system. Thus, the corresponding circuit can be modified to adapt its functionality to perform different tasks. Field Programmable Gate Arrays (FPGAs) are defined as semiconductor devices that are based around a matrix of configurable logic blocks (CLBs) connected via programmable interconnects. FPGAs can be reprogrammed to desired application or functionality requirements after manufacturing. This feature distinguishes FPGAs from Application Specific Integrated Circuits (ASICs), which are custom manufactured for specific design tasks;

The decision about the functional blocks to be implemented in software or in hardware in the dynamically reconfigurable embedded systems is based on the experience of the designer and/or making a brief exploration of the design space. This procedure, in addition to not complying with any methodology, does not ensure an optimal result, since for obtaining the best configuration it is necessary to solve an optimization problem which in most of its formulations is NP-hard [1]. In most cases, the HW/SW problem

DOI: 10.14738/tmlai.54.3205
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3205

is driven by optimization aim such as the area of hardware, execution time or power consumption. On the other hand, some of these models use exact algorithms to obtain an exact solution to the problem, but the search time increases proportionally to the problem size. Taking into account that in some cases a near optimal solution is considered as good enough [2]. Consequently, many researchers employed heuristics based on Genetic Algorithms [3][4][5], Simulated Annealing [6], Tabu Search [7], Ant Colony Optimization (ACO) [8] and bee swarms [9] to obtain sub-optimal solutions. The IA technics

are also used to resolve the HW/SW partitioning and scheduling problem. In [10], the authors apply a Fuzzy Logic for Hardware/Software Partitioning in Embedded Systems. A Neural Networks Based Approach for the Scheduling of Reconfigurable Embedded Systems supporting FPGA as reconfigurable component is applied in [11]. However there are a problems which are resolved by the hybridization of some of these algorithms and they have given an optimal results. For the resolution of HW/SW partitioning and scheduling problem is not yet adopted [12, 13].

Scheduling is the process of determining a starting time for each software/hardware task of the system. When it used with partitioning in codesign, it is also allows to determine the execution time of the system under design when a partitioning solution is determined (mapping of the software and hardware implementation tasks on the architecture target supporting a FPGA). The two problems partitioning and scheduling are interrelated and tightly coupled. Each of one is known to be NP-Hard, and the fact that they have to be solved simultaneously complicates their resolution. This is why the only feasible way to solve them is to use heuristics,since exhaustive methods take a prohibitive execution time to find the best solution as soon as the system becomes large [9].

This work describes an automatic partitioning/scheduling approach that uses the combination of global search represented by HBMO as main strategy with two local searches methods: HW1 and HW2 in order to intensify the search in promising zones detected by the HBMO exploration process. The HW1 and HW2 approaches are based respectively on Tabu Search (TS) and Simulated annealing (SA). These algorithms are used as heuristics in the stage of the brood's improvement HBMO. The comparison with the genetic algorithm (GA) has been used.

The paper proceeds as follows. The presentation and formulation of HW/SW partitioning and scheduling problem is described in Section II. Section III presents the proposed Scheduling Approach. The resolution process of the HW/SW partitioning and scheduling problem is detailed in Section IV. The V section, the simulation and experimental results are given. The paper finishes with the conclusion presented in the section VI.

2 HW/SW Partitioning and Scheduling Problem Formulation.

Our algorithm target architecture contains one processor, one dynamically reconfigurable devise of type FPGA, one bus and one shared memory. The bus is used as communication channel connecting processor and configurable devise.

The main goal of partitioning is to select whether to put each task of the application model into SW (task assigned to the processor) or HW (task assigned to the FPGA) such that the whole execution time is minimized and the area occupation is maximized while reducing the overhead. Each application task is defined as a 5-tuple: Ti = (wi, hi, sti, ci, hti) where wi, hi, sti, ci and hti, denote width, height, software execution time, communication cost and hardware execution time of the task. The reconfiguration time

of the task can be considered as a part of its hardware execution time (hti) . In this work, we focus on dependent tasks witch may be represented as a directed task graph (DAG).

In this model, a solution to the partitioning and scheduling is expressed as a set of binary elements: X={x_1, x_2,...,x_n} where the x_i element represents the type of implementation assigned to the task Ti; if x_i = 1, it means implementation on HW, or implementation on SW otherwise.

Basin of the given the above stated, it is possible to define the main objective taking into consideration all the aforementioned imposed constraints as following:

Let $T = \{T1, T2, .. Tn\}$ with i=1 to n

$$x_i = \begin{cases} 1 & if\ \ Ti\ \ is\ assigned\ on\ FPGA \\ 0 & if\ Ti\ is\ assigned\ on\ processor \end{cases} \quad (1)$$

$$Texecution = \sum_{i=1}^{N}[(1 - x_i) * sti + x_i * hti] + \sum_{i=1}^{N-1} c_i \quad (2)$$

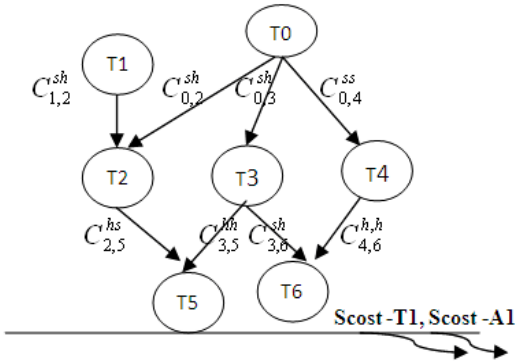
-c_i^{ss} (c_i^{hh}) denotes the communication time between node i and node (i+1) if both tasks blocks are assigned to software (hardware). $1 < i < n$

c_i^{sh} $\left(c_i^{hs}\right)$ denotes the communication time between node i and node (i+1) if task i is assigned to software (hardware) and task i+1 is assigned to hardware (software). 1 < i < n.

In our study the cost c_i^{ss} is ignored because two tasks i and i+1 are executed on the same processor. The total cost communication becomes:

$$c_i = x_i * x_{i+1} * c_i^{hh} + x_i * (1 - x_{i+1}) * c_i^{hs} + (1 - x_i) * x_{i+1} * c_i^{sh} \quad (3)$$

The following figure (fig.1) describes the tasks model used.



Bee-scheduling$_1$ [T_0, T_2, T_1, T_4, T_3, T_6, T_5]234 24
Bee-partitionning1 [hw_0,hw_2,sw_1,sw_4,hw_3,hw_6,sw_5]
Bee-scheduling$_2$ [T_0, T_2, T_4, T_1, T_3, T_5,T_6] 200 30
Bee-partitionning2 [sw_0, hw_2, hw_4,sw_1, sw_3,hw_5,hw_6]
Bee-scheduling3 [T_2, T_0, T_4, T_1, T_3, T_5 , T_6]322 9
Bee-partitionning3 [hw_2,sw_0,sw_4,sw_1,sw_3, sw_5,hw_6]
Bee-scheduling$_4$ [T_0 ,T_2 , T_1 , T_4 , T_3 , T_6 , T_5] 230 24
Bee-partitionning4 [hw_0,sw_2,sw_1,hw_4,hw_3,sw_6,sw_5]

Figure 1 HW/SW task model considered.Example of HW/SW partitioning-scheduling with cost function values.

In the third section, main aspects of our algorithm are described.

3 Proposed Hardware/Software Partitioning Scheduling Approach

The proposed approach combines the Honey Bees Mating Optimization (HBMO) algorithm with the local searches: WH1 based on simulated annealing (SA) and WH2 based on Tabu Search (TS). HBMO was found to outperform some better known algorithms. However, it has not been applied to dynamically reconfigurable embedded systems system scheduling/partitioning HW/SW problems. A honey bee colony consists of the queen(s), drones, worker(s) and broods. The Honey bees Mating Optimization algorithm mimics the natural mating behavior of the queen bee when she leaves the hive to mate with drones in the air. After each mating, the genetic pool of the queen is enhanced by adding sperm to her sperm theca. In [14], the authors give more information for the artificial comportment of the mating honey bees process. The workers are presented as heuristics whose functionality is to improve the broods produced during the mating process. The initial HBMO population is generated randomly and selects the best solution as the queen. To improve the broods produced during the queen's mating, we use two workers WH1 and WH2 based on local searches. The hybrid approach is based on a Honey Bees Mating Optimization (HBMO) that realizes a design space exploration by generating different mappings of the tasks on the processor and the FPGA.

4 Presentation of the Resolution Process

In this section, we present the approach adapted to resolve the issues addressed. To start with, a set of terms are explained in the table 1.

Table 1 Analogy between the natural bees, the artificial bees and the adopted partitioning/scheduling problem

Naturel honey bee entities	Artificial honey bee Entities	Scheduling/partitioning Entities problem (adopted approach)
Bees population	All feasible solutions	scheduling plans of hardware and software tasks with temporal and spatial constraints,
Queen	Best and optimal solution	Best HW/SW scheduling plan in the population
Drones	Incumbent solutions	Remaining HW/SW tasks scheduling plans.
Broods	New trial solutions	Plans gradually improved by crossover process in the mating HBMO stage and by two workers WH1 and WH2 in the HBMO amelioration stage.

The following algorithm presents the improved HBMO approach adopted.

Algorithm: HBMO_HW/SW partitioning scheduling

1. Initialization
2. **Generate the initial population of the honey bees randomly**
3. **Evaluate** the fitness of the scheduling (time execution cost, area cost);
4. **Select** the best plan of HW/SW partitioning of the population of HW/SW partitioning plans which represents the queen;
5. Sperm: Size of the spermatheca**;**
6. **E (t)** and **S (t)**: Energy and Speed in [0.5, 1];
7. **α**: factor of reduction of energy and speed in [0,1];
8. **M**: maximum number of mating flights;
9. **For** i:=0 **to** M (mating flights)
10. **do while** E(t) > 0 and Sperm is not full Select a drone
11. **If** the drone passes the probabilistic condition
12. **Add** sperm (plan) of the drone in the spermatheca
13. **Endif**
14. S(t+1) = **α** *S(t)
15. E(t+1) = **α** * E(t)
16. **Enddo**
17. **For** j = 1 **to** Size of Spermatheca
18. **Select** a sperm from the spermatheca;
19. **Generate** a brood by crossover the queen's genotype with the selected sperm;
20. **Improve the brood's fitness** by applying the **workers (WH1 &WH2)**
21. **If** the brood's fitness is better than the queen's fitness
22. **Replace** the queen with the brood
23. **else** Add the brood to the population of drones
24. **Endif**
25. **Replace** the drones by the broods
26. **Enddo(for)**
27. **Enddo**
28. **Return** The Queen (Best Solution Found).

Broods improvement: The improvement process is realized by two methods: WH1 and WH2.The genotypes resulting from the queen with the drone are improved by workers through WH1 and WH2. This stage of improved HBMO generates broods solutions: HW/SW partitioning. Thus the brood's improvement algorithm is applied to every operation of crosser often queen with the drone.

The next algorithms explain respectively the improvement strategies for the brood solutions by using two heuristic methods (Worker and WH2).

Algorithm WH1 (Worker Heuristic1:SA_ improving broods)

1. **Begin**
2. **Initial solution** s(T,A) , (brood which will be improved);
3. **Put** T1 ← cost-T(s(T))
4. T2←cost-A(s(A)) ; //do one mutation on the brood or One deplacement on the brood
5. s'(T,A)← **Mutation** (s(T,A));
6. **If** (cost-T(s'(T)) < cost-T(s(T))&
7. Cost-A(s'(A))>Cost-A(s(A)) **then** return s'(T,A)
8. **Else**
9. **Generate** r and r' randomly r in [0, 1]
10. **if** $\begin{pmatrix} r < e^{((cost-T(s(T))-cost-T(s'(T)))/T1} \text{and} \\ r' < e^{cost-A(s(T))-cost-A(s'(T))} \end{pmatrix}$
11. **then** return $s'(T, A)$
12. **Else** return s (T,A);
13. **Endif**;
14. **Endif**;
15. **End.**

Algorithm WH2 (Worker Heuristic1:TS_ improving broods)

1. **Begin**
2. Initial solution s (T,A) (brood which will be improved);
3. Insert **s** in the set L of solution ;
4. Scost-T_{min} ← s(T). / /Scost-T_{min} : best solution
5. Scost-A_{max} ← s(A)
6. **While** (Criterion of stop not checked) **do**
7. **Generate** the neighborhood of the current solution by **mutation**
8. **Select** s'(T,A) in this neighborhood although s'(T,A**)** is not present in the Set L.
9. **If** Scost-T(s'(T)) < Scost-T_{min}(s(T)) and Scost-A(s'(A)) > Scost-A_{max}(s(A)) **then begin**
10. Scost-T_{min} ← s'(T)
11. Scost-A_{max} ← s'(A**)** //minimize the cost function time and maximize cost function area
12. **Update** the set L;
13. **End;**
14. **Endif;**
15. **End of while;**
16. **End.**

5 Simulation and Experimental Results

In this section results of the Improved HBMO for HW/ SW partitioning and scheduling are compared to those given by the genetic algorithm (GA). In GA each task is represented by a binary gene which denotes the allocation of this task in the physical resources. Another advantage of this coding strategy is that genetic operators, such as crossover and mutation, can easily be applied without generating invalid chromosomes. It is also easy to extend to loosely coupled multiple-processor mulitple- FPGA architecture.

The JCreator version 4 has been used to implement the approach. These generated graphs are used as the benchmark to evaluate the result quality of our HW/SW partitioning and scheduling. In the graph, Each node i is represented by the quadruplet (sti,hti,ci, hai(wi,hi)) contains software execution time (sti), hardware execution time (hti), cost communication (ci) and hardware area occupied (hai)such the area hai is defined by the product between the width (wi) and the height (hi). It specifies the number of CLBs necessary for the execution of the task in FPGA. The values of software execution time (sti), hardware area occupied (hai) and hardware execution time (hti) were generated randomly. Since the execution time for one node implemented in software is generally greater than in a hardware implementation. The execution time of hardware (FPGA) equals one third to one fifth of the execution time of software [15].

5.1 Parameters values

The parameters of the adopted approach have been selected after thorough testing. A number of different values were tested and the ones selected are those that yielded the best results in both solution quality and computational time. The best values of these parameters are presented in the following tables:

Table 2. HBMO and GA Parameters values

Algorithms	Parameters	Values
HBMO	Population size	50
	Mating flight	100
	Spermatica	6
	Speed	0.80
	Energy	0.70
	Reduction factor α	0.20
Genetic Algorithm	Population size	50
	Probability of crossing	0.60
	Probability of mutation	0.20
	Number of generation	70

Table 3. Workers_ Parameters values

Algorithms	Parameters	Values
Worker_ simulated annealing (WH1)	Initial temperature	Brood-cost
	Final temperature	0
	Number of iterations	Until the amelioration or temperature=0
Worker_ Tabu search (WH2)	Size of list	5
	Number of iterations	5
	neighbood	Mutation and deplacement

5.2 Simulation results

In this section, we study the effectiveness of HBMO approach. The comparisons were made on the values found by the application of each algorithm: Improved HBMO and GA. In each execution evaluations of both objective functions were made: minimization of execution time and maximization of occupation area. This is illustrated by the graphs presented in the Fig.1 and Fig. 2:

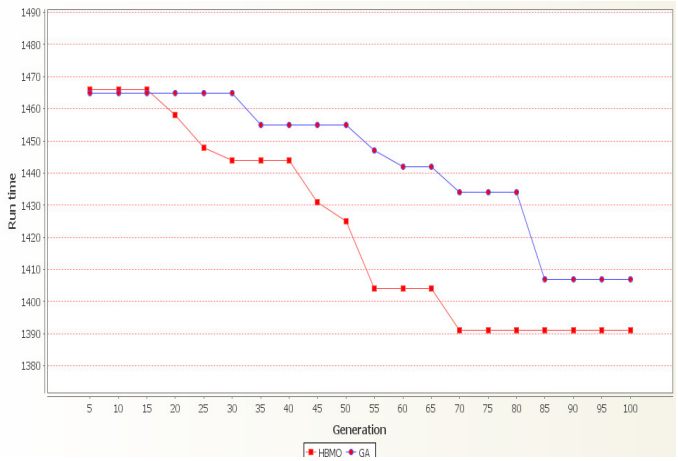


Figure 2. Comparative results between the graphics simulations HBMO (red curve) approach and genetic algorithms GA (blue curve) (case of minimization of execution time)

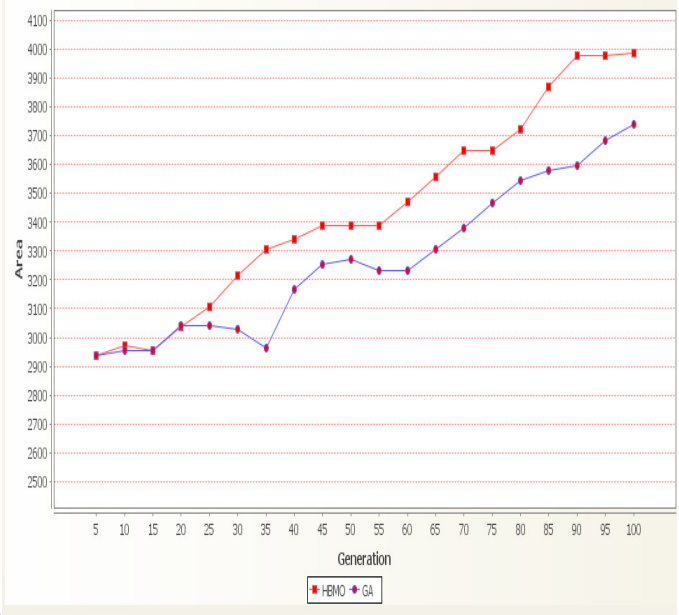


Figure 3 Comparative results between the graphics simulations HBMO approach (red curve) and genetic algorithms GA(blue curve) (case of maximizing the space used for FPGA size = 2000 CLBs.

We note that the effectiveness of HBMO is proven when ever the size of the FPGA increases. The majority of tasks will have hardware implementations. The same applies to the execution time, we note that HBMO converges to a minimum execution time than the GA.

6 Conclusion

In this paper, we described an improved HBMO heuristic for scheduling and HW/SW partitioning applications on reconfigurable System on Chip (RSoC). The application of the adopted method shows that the two problems: HW/SW partitioning and scheduling are hence interrelated and tightly coupled because the partitioning is included in the design space exploration while scheduling allows the evaluation of each solution. To test the validity of these algorithms, the results are obtained by HW/SW partitioning and

scheduling different groups of task graphs on different target systems (using different number of resources: size of FPGA). We compared the adopted approach, HBMO, to GA method, the simulation results obtained solutions showed the optimality for scheduling process and the effectiveness of the adopted approach in the all cases of examples.

REFERENCES

[1] Y. Jing, J. Krang,Jiayi DU and B. HU, Application of improved simulated Annealing optimization Algorithms in hardware/Software partitioning of the system on chip,Sringer.pp532-540,2014.

[2] H. Daz Pando, S. Asensi, R. Seplveda et al. An Application of Fuzzy Logic for Hardware/Software Partitioning Embedded Systems. Computation system Vol. 17 No.1, pp. 25-39.2013.

[3] M.B. Abdelhalim, S.E.-D.Habib,An integrated high-levelhardware/software partitioning methodology, Des Autom Embed Syst 15: pp.(19-50), 2011.

[4] P. Arat,., Z.A. Mann, and A. Orbn. Algorithmic aspects of hardware/software partitioning. ACM Transactions on Design Automation of Electronic Systems (TODAES), 10(1), 136156. 2005.

[5] S. Luo, X. Ma, Y. Lu .: An Advanced Non-Dominated Sorting Genetic Algorithm Based SOC Hardware/Software Partitioning, ACTA ELECTRONICA SINICA, (11):pp. 2595-2599. 2009.

[6] Y. Kang, H. Lu, J. He.: A PSO-based Genetic Algorithm for Scheduling of Tasks in a Heterogeneous Distributed System, Journal of software, 8(6): pp.1443-1450. 2013.

[7] F.Vahid Modifying min-cut for hardware and software functional partitioning. 5th International Workshop on Hardware/Software Co- Design (CODES/CASHE97), Braunschweig, Germany, 4348 .1997.

[8] F. Ferrandi, P. C. Lanzi, C. Pilato, D. Sciuto, A. Tumeo. Ant Colony Optimization for Mapping, Scheduling And Placing in Reconfigurable Systems. NASA/ESA Conference on Adaptive Hardware and Systems (AHS),2013.

[9] M. Koudil , K. Benatchba, ATarabet, E. B. Sahraoui. Using bees to solve partitioning and scheduling problem in codesign . Applied mathematics and computation 186,pp1710-1722. 2007.

[10] H. Daz Pando1, S. Cuenca Asensi, R. Seplveda Lima, J. Fajardo Caldern1 and A. Rosete Surez, An Application of Fuzzy Logic for Hardware/Software Partitioning in Embedded Systems. Computacin y Sistemas Vol. 17 No.1, pp. 25-39, 2013.

[11] G. Rehaiem , H. Gharsellaoui, S. Ben Ahmed. A Neural Networks Based Approach for the Real-Time Scheduling of Reconfigurable Embedded Systems with Minimization of Power Consumption. ICIS 2016, Okayama, Japan June 26-29, 2016.

[12] P.Peng, Z. Kuchcinski, K.Doboli, A.. System level hardware/ software partitioning based on simulated annealing and Tabu Search. Design Automation for Embedded Systems, 2(1), 532. 1997.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Hybridized Model for Early Detection and Smart Monitoring of Forest Fire

Mohammed Anas El abbassi, Abdelilah Jilbab, Abdennaser Bourouhou
Electrical Engineering Research Laboratory
ENSET-Rabat, Mohammed V University in Rabat
Rabat, Morocco
elabbassi.3@gmail.com

ABSTRACT

The demand for wireless sensor network technology has been increasingly needed in recent years for several major applications, including environmental monitoring, where nodes deployed in nature detect, process and transfer the environmental data in an autonomous way. However, the performance of early detection of the fire event after data fusion processed by this WSN's system will be less reliable since the majority of the other nodes do not detect the fire yet (at the beginning of event). That is why the present work is conducted. It proposes an intelligent strategy of data fusion of temperature and humidity sensors hybridized with an intelligent scanning technique. This model will allow the early detection of alarm from the beginning of fire event and guarantee a good monitoring of area state with an ongoing localization of fire zone. The result proves a very good performance in terms of reliability of the early detection and tracking fire propagation.

Keywords— Wireless sensor networks (WSN); early detection; data fusion; scanning technique; fire event; monitoring.

1 Introduction

As the exploitation of the wireless sensor networks [1] is strongly demanded in several fields, such as the environment (as discussed in [2]-[3]-[5]-[12]), the army and medicine...etc, some scientific researchers still lay their focus on the same direction in order to find promising resolutions to improve the performance of these sensors in terms of energy consumption, routing protocols, compression and transmission of distributed data, and collaborative signal processing which shapes our main interest, particularly, the processing of the multi-sensor data fusion.

Regarding the studies on the reliability of alerts, it is affirmed in several related works that the fusion of information multi-sensors plays a primordial role for a good reliability of alert. However, several fusion models (as in [4] - [5]) treat the alert issue based on the acquisition of data provided from the nodes deployed in the surveillance field, in the case of a fire event and at the very beginning, minorities of the nodes immediately detect the fire, while, the other nodes that are still far from the event, have not detected the fire yet. The system, then, starts the calculation based on all the collected data. Such reasoning requires sufficient propagation of the fire before it is detected by the system, which

DOI: 10.14738/tmlai.54.3206
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3206

systematically generates an alert delay that can cause burns of large areas of the monitored environmental zone.

In addition, many of these proposed approaches of researchers (as in [4] and [5]), have not discussed the information filtering processing, which is an interesting operation that allows to avoid the impact of erroneous data on the processing of information fusion.

It is in this regard, the present work proposes an intelligent architecture of data fusion related with a scanning technique in order to reliably detect and locate the fire zone at the beginning of its propagation in the area. Such a model also takes into account the intelligent elimination of erroneous collected information.

2 Proposed Work

2.1 Global Model Architecture

For the discussion of the proposed model, let's consider that n nodes are dispersed in the monitored environment; these micro-sensors are capable to detect, collect and transfer the collected environmental data in an autonomous way (physical quantity of temperature and humidity) to the Sink point through a specific routing of the data. This Sink processes these data and transfers them to the control center where the central computer performs processing, fusing and decision-making operations.

This part intends to present a global fusion plan. This model reacts intelligently from the beginning of the fire with a minimum possible of propagation. Generally speaking, this model discusses a two data fusion modes hybridized with a scanning algorithm applied to the matrix of collected measurements. This proposed model contains global hierarchical steps (Fig. 1), each of which will be discussed separately in the present work.

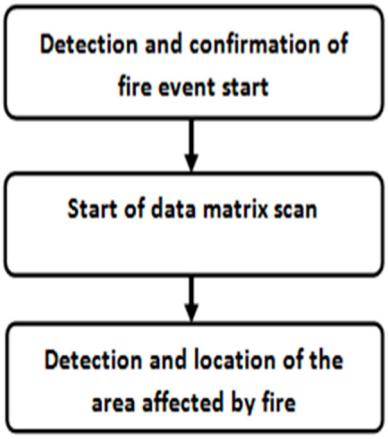


Figure 1 Global model process.

2.2 Description of The Steps

2.2.1 Primary Fire Detection

This part discusses a fusion approach that guarantees detection of fire from its first beginning by using at first the *KNN* classifier [6]-[7] for making a local decision in single node (first mode of data fusion), also, allowing to make a final reasoning of fire event based on collected samples of the neighboring nodes by

using the Equal Gain Combiner (*EGC*) fusion rule [8] chained to the *KNN* classifier (second mode of data fusion).

a) First Mode of Data Fusion

This part focuses on the application of *K*-Nearest-Neighbors algorithm (*KNN*) [6]-[7] for evaluation of the combined data resulting from temperature and humidity measurements (see Fig.3) of single node. And to define the *KNN* algorithm, the latter is among the most common methods for classification of the objects, applied since 1970's as a statistical estimation technique, this classification is based on a voting system of its *K* nearest neighbors which are measured by the similarity function (or distance) relative to the target data [6]. This distance is usually measured by applying the Minkowski distance [9]-[10], described as following:

$$MD = \left(\sum_{l=1}^{u} (|x_l - y_l|)^p\right)^{\frac{1}{p}} \quad (1)$$

Where x_l and y_l are the coordinates of the two points *X* and *Y*, while the parameters: *u* and *p* are selected by user (*u*=2, *p*=2 in our approach).

Let *n* = 100 nodes geographically dispersed in a surveillance zone with a surface of 90x90 m^2 where each known location node is distanced by 10m from its nearest neighbors. These nodes are capable to measure the physical quantities of temperature and humidity.

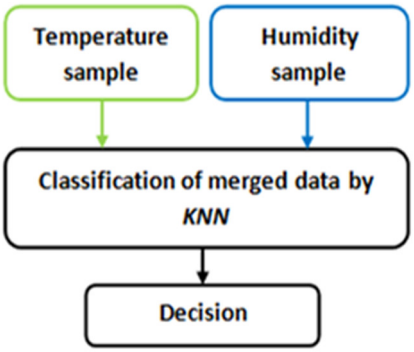


Figure 2 Local fusion and decision process inside node (First mode of data fusion).

Therefore, to apply this local data fusion operation by using *KNN* (as shown in Fig.2), we have in the same node C_x (x is an index, x ∈ [1; 100]), a case of collection of two physical quantities samples of temperature type denoted Ts_x, and of humidity type denoted Hs_x. Given that $D_{Local_x}(Hs_x, Ts_x)$ represents a result point based on these two coordinate samples where each combined $D_{Local_x}(Hs_x, Ts_x)$ must be processed and assigned to one of two clusters : {Fire Cluster} and {None Fire Cluster} as described in Fig.3.

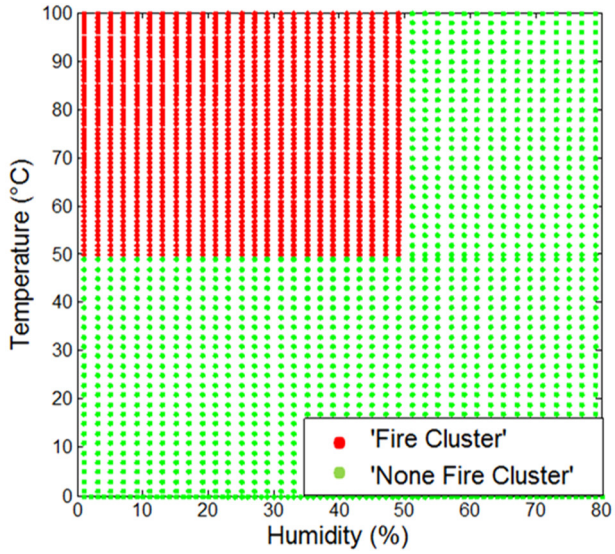


Figure 3 Definition of clusters regions according to temperature and humidity ranges.

As you can see, the previous clusters which are defined on two axes (Humidity and Temperature), express opposite hypotheses, their intervals in both axes, are carefully chosen by the user, taking into account the favorable conditions that generate the fire and also the climate of this geographical area being monitored. Note that these clusters are stored on the database and used as basic information on which *KNN* classifier can refer to decide.

According to Fig.3, the D_{Local_X} point expresses the danger state if it's assigned to {Fire Cluster}. Until now, as mentioned, it is information considered independent and specific to a single node. And to avoid any errors or malfunction or false testimony from a failing node, as a node detects such an alarm, the system reacts and begins to seek confirmation of the information by the occurrence of the nearest neighboring node (Fig. 4). At this stage, an analysis of fusion operation of these samples including the sample of the first sensor is initiated in order to confirm the presence or none of the fire, this new operation of data fusion called a second mode of data fusion that will be discussed in the next section.

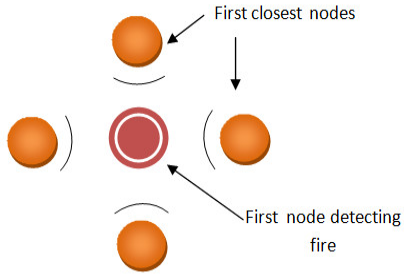


Figure 4 Example of the first node with its four neighboring nodes detecting the fire (*N*=5)

b) Second Mode of Data Fusion

Let *MD* (2), a matrix of the local estimation samples denoted Dp_i (i =1, 2…n) provided by n = 100 nodes, where Dp=1 represents a 'fire state' while Dp =0 represents a 'none fire state', those decisions are collected for each t=τ.

$$MD = \begin{bmatrix} Dp_1 & \cdots & Dp_{10} \\ \vdots & \ddots & \vdots \\ Dp_{91} & \cdots & Dp_{100} \end{bmatrix} \quad (2)$$

We take the case of detection of the first node C_x located inside the matrix, which means that Dp_x=1 as already mentioned. In this respect, the data fusion based on this sensor and the four surrounding neighbors is established to testify about the existence of fire or none. The proposed processing of the second mode of data fusion has been broken into phases which have been illustrated in the following Fig.5

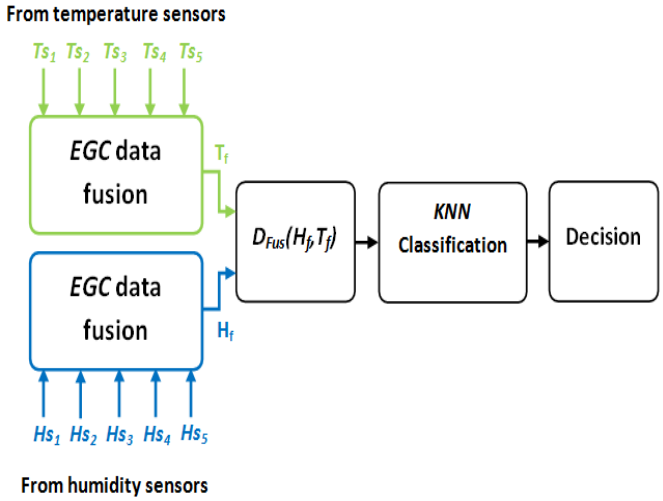


Figure 5 The proposed process of the second mode of data fusion over five nodes.

The first stage processes the combination of measures shipped from sensors of the same category by applying the *EGC* fusion rule [8], the latter is among the most common algorithms of data fusion, it has thus shown its effectiveness in terms of speed in processing and its reliability of output with a minimum amount of information required, more than that, compared to some algorithms as mentioned in [8], it is considered robust and stable for a wide range of *SNR*.

At the output of the *EGC* blocks, we find ourselves with two fused samples of temperature and humidity denoted T_{fx} and H_{fx}, the latter are subsequently the coordinates of a combined sample denoted $D_{Fusx}(H_{fx}, T_{fx})$, which is acted as a target data to the *KNN* algorithm, this one, classifies the new data given the same reasoning previously discussed in the local fusion with a slight change in the boundaries of the intervals of the two clusters, this change in limits is taken into account in this mode of fusion in order to guarantee the correct decision based on a majority of the nodes among N (N_{max}=5) which provides measurements belonging in the same hypothesis. The output of the processing translates the final decision noted F_{Dx}, which estimates the state of the area covered by these nodes.

Parallel processing begins on the matrix zone (Fig.6) after detecting the fire by the first node C_x, as already mentioned, the system takes the state of solidification of the information by application of data fusion with the neighboring nodes, this algorithm takes into account the location of the first node detecting the alarm with their neighbors situated in this matrix zone, consequently, the number *N* of the nodes varies automatically between 3 and 5 as described in Fig.6.

According to Fig. 6, the primary analysis fusion can be established on the basis of a number *N* of nodes that are located in one of the 3 mentioned regions. The first region, which is presented inside the matrix, processes the data fusion of the first node with its four neighbors. The second region comprises the four matrix edges (without corners) where the data fusion of four nodes is initiated. The third region includes the four corners of the area with an established data fusion of three nodes.

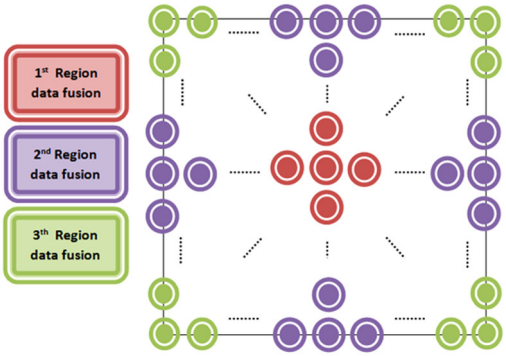


Figure 6 Definition of three different regions of data fusion in the matrix zone.

2.2.2 Scanning of The Matrix Data

After confirming a beginning of a fire start by the proposed technique above, it becomes essential to disclose continually the areas affected during the propagation of the fire. The scanning operation of the matrix zone is performed from the first affected area by a reasoning of data fusion of *N* nodes, the same operation is conducted again after incrementing or decrementing the scanning process in the matrix.

The scan, in this case, divides the matrix from the first detection point into two zones namely: A and B as shown in Fig. 7 (it may be just a single zone of the entire matrix if the fire starts from one of the four regions corners). Thus, this scan starts from the same detecting point and scans these two zones to the end points taking into account that each new data fusion is processed based on a number *N* (3, 4, or 5 nodes) depending on the location of the fusion operation on the matrix during scanning.

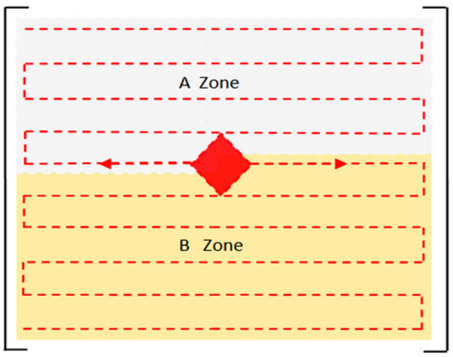


Figure 7 Scanning of the data matrix in two zones A and B (case of first detection inside the matrix zone).

On the one hand, this operation contributes to identifying the sensors that detect the fire and to avoiding sensors providing incorrect data or sensors that are located in an area that is still intact. On the other hand, it allows determining the direction and the extent of the propagation of the fire and also monitoring the surface of the burned zone. This reasoning will be described in the next section.

2.2.3 Detection, Location and Monitoring of The Affected Area.

At the end of the scan, a new matrix of final binary decisions denoted M_{ext} (3) is extracted. Through this matrix, we can actually illustrate the state of each part of the zone. Obviously, this illustration comes after having decided (fire or none-fire) based on each point of M_{ext}, where each point translates the state of the local zone under the surveillance field of N neighboring nodes after data fusion. Knowing that the system already notes the location of each sensor node, it is easy to identify in this case the zones affected by the fire.

$$M_{ext} = \begin{bmatrix} F_{D1} & \cdots & F_{D10} \\ \vdots & \ddots & \vdots \\ F_{D91} & \cdots & F_{D100} \end{bmatrix} \quad (3)$$

In order to trace the direction of propagation of the fire, the centroid that is based on the nodes responsible for the positive decisions (fire event), is calculated at each new iteration of the scan. This centroid is updated at each scanning operation, as shown in Fig.7, noting that each calculated centroid becomes the starting point for the new scan processing.

2.3 Simulation Results

Here is an example of a simulation of fire propagation in a square matrix surface monitored by 100 nodes, after detecting the first fire event. The system begins scanning iterations and makes a continuous update of the affected area with tracing fire propagation path, as illustrated in Fig.8.

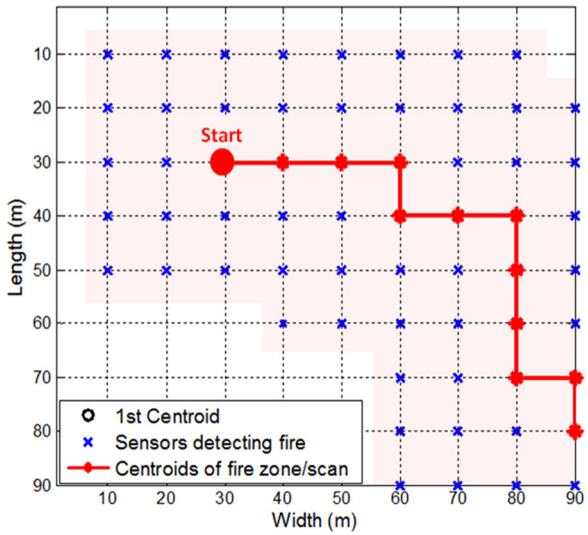


Figure 8 Monitoring of the affected area with propagation path tracing. (Example of area burning under 62 nodes detecting fire, while the area monitored by other 38 nodes is still intact).

And, to evaluate the alarm response time, let's consider another case where the fire propagates on a surface of a circle the center of which is located by the first detector node and has a radius that increases by 1 m/s. The response of alarm detection will be illustrated in Fig.9.

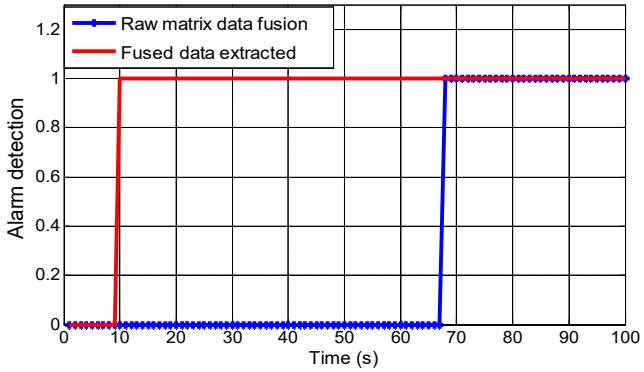


Figure 9 Responses comparison of alarm detection between the proposed model of data fusion and the raw data fusion of the entire matrix samples.

Fig.9 shows the positive response of the proposed method (in red line), where the alarm is triggered after a period of 10s, which coincides with a radius of 10m. The latter is reasoned by the distance fixed between the first detector and its neighbors. On the other hand, the alarm of a data fusion response of the matrix of the raw samples (in blue line) is triggered after a delay time of 67s which coincides with 67m of the radius of the burned zone circle; therefore, it can really causes great environmental damage. On the premise of this reasoning, the proposed method shows its efficiency in terms of vigilance and rapidity of alert response.

3 Conclusion

In this work, we have presented a new approach of multi-sensor data fusion with two types of physical quantities (temperature and humidity) integrated in nodes deployed in a surveillance environmental area. It has also presented an intelligent strategy of hybridization between a data fusion operation and a scanning technique. In brief, this approach aims at making the early detection of fire event more reliable, with a minimum possible of fire damage while maintaining monitoring of the area state and defining a continuous tracing the fire propagation path.

The simulation results of the proposed approach show a very effective performance of early detection and prove efficient monitoring and accurate localization of the affected area.

We have considered using fast response algorithms (such as the *KNN* classifier and the *EGC* fusion rule), which are based on a simple arithmetic calculation that allowed for a flexible processing with a low energy consumption in WSN's system.

REFERENCES

[1] K. Sohraby, D. Minoli, T. Znati ." Wireless sensor networks: technology, protocols, and applications", John Wiley & Sons (2007).

[2] M. Hefeeda,M. Bagheri,"Forest Fire Modeling and Early Detection using Wireless Sensor Networks", Ad Hoc & Sensor Wireless Networks Vol. 7, pp. 169–224,2009.

[3] E. Sisinni, A. Depari, A. Flammini," Design and implementation of a wireless sensor network for temperature sensing in hostile environments", Sensors and Actuators A: Physical, vol.237, pp. 47-55, 1 January 2016.

[4] E. Zervas , A. Mpimpoudis, C. Anagnostopoulos , O. Sekkas, and S. Hadjiefthymiades,"Multisensor data fusion for fire detection", Information Fusion ,vol.12 :150–159, 2011.

[5] M .A.El abbassi, A.Jilbab,A.Bourouhou,"A robust model of multi-sensor data fusion applied in wireless sensor networks for fire detection",International Review on Modelling and Simulation (IREMOS), vol.9, N°3,pp.173-180, Vol. 9, N. 3,pp 173-180, June 2016.

[6] Massart, D.L. "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules." Anal. Chim.Acta 136, 15-27, January 1982.

[7] Vijay Kotu, Bala Deshpande, "Predictive Analytics and Data Mining," ,chapitre 4, Pages 63-163 , 2015.

[8] R. Niu ,B. Chen, P.K. Varshney," Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks ",IEEE Transactions on signal processing,Vol.54,NO.3,March 2006.

[9] Teknomo, Kardi(2015), "Similarity Measurement" http://people.revoledu.com/kardi/tutorial/Similarity/MinkowskiDistance.html ,(2015).

[10] M .A.El abbassi, A.Jilbab,A.Bourouhou, " Detection model based on multi-sensor data for early fire prevention ", IEEE ICEIT tangier Morocco, pp.214 – 218. 2016.

[11] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, "Multisensor data fusion :A review of state-of-the-art, Information fusion", vol 14: 28-44,2013.

[12] M.A.El abbassi, A.Jilbab, A.Bourouhou," Fusion des données dans les réseaux de capteurs sans fils pour la protection de l'environnement ", TELECOM'2015 & 9ème JFMMA, Meknès- Maroc, 2015.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Role of Management and Policy Issues in Computer Security: Rand Report R-609 within Organization

N. Tagmouti, Jamal Boujmil, N. Raissouni
RS&GIS Lab. Dept. Telecommunications
The National School for Applied Sciences of Tetuan
Tetuan, Morocco
jamal.boujmil@gmail.com, nohatg@gmail.com, nraissouni@uae.ma

ABSTRACT

The need to provide strengthened Security for Information Systems within organization increases day after day seeing the large development of interconnection of the World Wide Web and the clear effect that results by the frequent and multiple productions of attacks and threats. This feebleness and gap in current information technologies urge researcher to face by developing more secure systems of Sharing Resources. The history of securing Information Systems began with the concept of Computer Security in its Global meaning Physical and Non-Physical. Thus, by 1967 the Department of Defense of USA published the R-609 which is considered as the first step in the wide world of Information security including Securing the data, Limiting random and unauthorized access to that data and Involving personnel from multiple levels of the organization in information security. The purpose of this paper is to analyze and evaluate the role of management and policy issues in computer security that proposes Rand Report R-609 within organization.

Keywords: Information Systems, Information Security, R-609

1 Introduction

Information security begins with computer security. The Advanced Research Projects Agency (ARPA), during June 1967, designed a special commission to carry out a study and research of the process of securing classified information systems. This commission assembled in October 1967 with regular meetings has formulated recommendations ultimately becoming the contents of the Rand Report R-609 [1].The need for computer security (i.e., the need to secure physical locations, hardware, and software from threats) ascended during World War II when the first mainframes, developed to aid computations for communication code breaking, were put to use.

The Rand Report R-609isconsidered the pioneer and formerextensively recognized issuedmanuscript to recognize the role of management and policy issues in computer security. It is noted in the report, that the wide exploitationin information systems of networking components is introducing security risks. These last not mitigated by the routine practices used then to secure these systems. The reportindicated a crucial moment in computer security historywhen the choice of computer security lengthenedmeaningfully from the safety of physical locations and hardware to take account of the following points:

DOI: 10.14738/tmlai.54.3209
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3209

- Securing the data
- Limiting random and unauthorized access to that data
- Involving personnel from multiple levels of the organization in matters pertaining to information security

Our objective in the present research is to carry out an analysis and a comparative study based on the documents dealing with role of management and policy issues in computer security. The study will focus mainly on the multiple levels of security been implemented to protect the mainframes and maintain the integrity of the data (e.g., access to sensitive military locations was controlled by means of badges, keys, and the facial recognition of authorized personnel by security guards). Furthermore, the growing need to maintain national security eventually led to more complex and more technologically sophisticated computer security safeguards.

2 Arpanet Program Plan Development

ARPANET come to bestandard and more widely used during the 1970s and 80s, and the potential for its misuse developed. In December of 1973, Robert M. “Bob” Metcalfe, creditedwith the development of Ethernet (one of the furthermostwidely held networking protocols) identified major problems with ARPANET security. Individual remote sites are not having sufficient controls and safeguards to protect data from unauthorized remote users. In addition, other difficultiesoverflowed [1]:

Password structure and formats vulnerability.

Lack of safety procedures for dial-up connections.

User identification and authorization to the system absence.

Phone numbers were commonly distributed and openly exposed on the walls of phone booths,providing hackers easy access to ARPANET. Due to these drawbacks, and because of the range and frequency of computer security violations and the explosion in the numbers of hosts and users on ARPANET, network security was referred to as network insecurity [2].Subsequently, in 1978, theeminentresearch entitled “Protection Analysis: Final Report” was appreciatively published. This study focused on the project started by ARPA to discern the weaknesses of operating system security. Chronologically, including this and other seminal studies in early computer security shown below:

1968- Maurice Wilkes discusses password security in Time-Sharing Computer Systems.

1973- Schell, Downey, and Popek examine the need for additional security in military systems in “Preliminary Notes on the Design of Secure Military Computer Systems”[3].

1975- The Federal Information Processing Standards (FIPS) examines Digital Encryption Standard (DES) in the Federal Register.

1978- Bisbey and Hollingworth publish their study “Protection Analysis: Final Report,” discussing the Protection Analysis project created by ARPA to better understand the vulnerabilities of operating system security and examine the possibility of automated vulnerability detection techniques in existing system software[4].

1979- Morris and Thompson author “Password Security: A Case History,” published in the Communications of the Association for Computing Machinery (ACM). The paper examines the history of a design for a password security scheme on a remotely accessed, time-sharing system.

1979- Dennis Ritchie publishes “On the Security of UNIX” and “Protection of Data File Contents,” discussing secure user IDs and secure group IDs, and the problems inherent in the systems.

1984- Grampp and Morris write “UNIX Operating System Security.” In this report, the authors examine four “important handles to computer security”: physical control of premises and computer facilities, management commitment to security objectives, education of employees, and administrative procedures aimed at increased security[5].

1984- Reeds and Weinberger publish “File Security and the UNIX System Crypt Command.” Their premise was: “No technique can be secure against wiretapping or its equivalent on the computer. Therefore no technique can be secure against the systems administrator or other privileged users … the naïve user has no chance” [6].

The programconcerning security that went beyond protecting physical locations began with a single paper sponsored by the Department of Defense, the Rand Report R-609. Whichattempted to define the multiple controls and mechanisms necessary for the protection of a multilevel computer system. The paper was confidential for almost a decade, and is nowadays, considered thedocument that impulses the study and analysis of computer security.

Therefore, the role of management and policy issues of security of the entire systems sharing resources inside the Department of Defense was brought to the attention of researchers in the spring-summer of 1967. Meanwhile, systems were being acquired at a rapid rate and securing them was a pressing concern for both the military and defense contractors.

3 Security Control Definitions

A wide terminologyis being used in relationtosecurity control for which practice is not completely standardized. In this paragraph, we present the terms used throughout the Rand Report R-609 respecting the way they were defined as a group [1-6].

• **CLEARANCE:** The privilege granted to an individual onthe basis of prescribed investigative procedures to have formal access to classified information when such access is necessary to his work. The three formal national clearances are Top Secret, Secret, and Confidential. However, it is also expedient from the computer point of view to recognize Uncleared as a fourth level of clearance. A clearance is a necessary but not sufficient condition to have access to classified information. By extension, the concept of clearance can be applied also to equipment. For example, when a computer terminal is spoken of as having a given level of clearance, it is implied that certain investigative procedures and tests have established that the corresponding level of classified information can be safely transmitted through that terminal. When referring to an aggregation of equipment, together with its management controls and procedures, facility clearance is some- times used.

• **NEED-TO-KNOW**: An administrative action certifying that a given individual requires access to specified classified information in order to perform his assigned duties. The combination of a clearance and a need-to-know constitutes the necessary and sufficient conditions for granting access to classified information.

• **CLASSIFICATION**:The act of identifying the sensitivity of' defense information by ascertaining the potential level of damage to the interests of the United States were the information to be divulged to an unfriendly foreign agent. The classification of in formation is formally defined in Executive Order 10501. There are only three formal levels of national classification: Top Secret, Secret, and Confidential, but it is expedient from the computer point of view also to consider Unclassified as a fourth level of classification. The identifiers associated with an item of classified information, indicating the level of classification or any special status, are generically called labels.

• **SPECIAL CATEGORY = SPECIAL-ACCESSCATEGORY =COMPARTMENT**:Classified defense information that is segregated and entrusted to a particular agency or organizational group for safeguarding. For example, that portion of defense classified information that concerns nuclear matters is entrusted to the Atomic Energy Commission, which is responsible for establishing and promulgating rules and regulations for safeguarding it and for controlling its dissemination. Classified information in a special category is normally identified by some special marking, label, or letter; e.g., AEC information, whether classified Confidential, Secret, or Top Secret, is collectively identified as Q-information. It is often called Q-classified, but note that this use of classification is an extended sense of the formal usage of the word.Sometimes, special investigative procedures are stipulated for granting access to information in special categories. Thus, while formally there are only three broadly defined national clearance levels, in practice there is a further structure within each level. In part, this reflects the separation of information into special categories, and, in part, the fact that many different agencies are authorized to grant clearances. For example, an individual functioning within the AEC domain and cleared to Top Secret will often be said to have a Q-clearance because he is authorized access to Top Secret information entrusted to the AEC for safeguarding and identified by the special category Q. These special types of clearances at given levels are not always specifically identified with a unique additional marking or label.

• **CAVEAT**: A special letter, word, phrase, sentence, marking, or combination thereof, which labels classified material as being in a special category and hence subject to additional access controls. Thus, a caveat is an indicator of a special subset of information within one or more levels of classification. The caveat may be juxtaposed with the classification label may appear by itself; or sometimes does not appear explicitly but is only inferred. Particular kinds of caveats are:

- CODEWORDS: An individual word or a group of words labelling a particular collection of classified information.
- DISSEMINATION LABELS = ACCESS CONTROL LABELS: A group of words that imposes an additional restriction on how classified information can be used, disseminated, or divulged; such labels are an additional means for controlling access. Examples: "No Foreign Dissemination," "U.S. Eyes Only," "Not Releasable Outside the Department of Defense".
- INFORMATION LABELS: A group of words that conveys to the recipient of information some additional guidance as to how the information may be further disseminated, controlled, transmitted, protected, or utilized. Examples: "Limited Distribution," "Special Handling Required," "Group 1 Excluded from Automatic Downgrading and Declassification".

• **FULLY CLEARED**: An individual who has the clearance and all need-to-know authorizations granting him access to all classified information contained in a computer system. By extension, the term can be applied to equipment, in which case it implies that all necessary safeguards are present to enable the equipment to store and process information with many levels of' classification and caveated in many different ways.

• **SECURITY FLAG**: For the purposes of this Report. It is convenient to introduce this new term. It is a composite term, reflecting the level of classification. All caveats (including codewords and labels), and need-to-know requirements, which together are the indicators establishing the access restrictions on information or the access privileges of an individual. By extension, the concept can be applied to equipment and indicates the class of information that can be stored and processed.Thus, the security flag contains all the information necessary to control access. One security flag is considered to be equal to or higher than a second if a requestor with the first flag is authorized access to information which has the second flag.

• **SECURITY PARAMETERS**: The totality of information about users, files, terminals, communications, etc., which a computer system requires in order to exercise security control over the information that it contains. Included are such things as user names, clearances, need-to-know authorizations, physical location; terminal locations and clearances; file classifications and dissemination restrictions. Thus, a set of security parameters particularizes a generalized security control system to the specific equipment configuration, class of information, class of users, etc., in a given installation.

4 Management and Administrative Control

An effective andagreed set of management and administrative controls and procedures governing the information'sstreamto and from the computer systemmust be added to overall policy guidance and to technical methods. In addition to the movement and actions within the system environment of people and transportable components. The Standardization of activities and the requirement for standards all through the system is the essential aspect of effective and agreed control.Standards are effective in severalapproaches. Thus, with strictlyagreed procedures, the different operators will be reticent from taking shortcuts that can result in leakage. Therefore, typical procedures that are required with some details of each are presented in the following[1-6].

4.1 Operational Start-Up

Procedures must be established for putting a resource-sharing system into operation, and must include provisions for loading a fresh, certified copy of the Supervisor software, for verification of its correct loading, for validation of system security checks, for inserting relevant security parameters, and for certification of system security status by the System Security Officer.

4.1.1 Scheduled shutdown

The procedures for a scheduled shutdown of operations must take account of proper notification of the System Security Officer, physical protection of demountable storage (tapes, discs) as required, orderly closing of internal files, validation of the suspension of operation of all terminals, demounting of all copies (or required parts) of the Supervisor software, erasure of any parts of the Supervisor software remaining in working storage, verification of erasure of the Supervisor, disconnection of remote communication circuits, and physical securing of the power controls.

4.2 Unscheduled shutdown

An unscheduled shutdown must initiate procedures for immediate surveillance and recording of all indicators .to help ascertain what happened; any needed emergency actions in case of fire, water hazard, etc.; special surveillance or physical protection measures to guarantee that no demountable items are removed; immediate notification of the System Security Officer; and special security controls (for example, protecting all printouts, including those at terminals, in accordance with protection rules for the highest classification handled in the system until the situation can be resolved).

4.3 Restart after unscheduled shutdown

If a trouble condition has caused the system to shut down, it is necessary that there be procedures to handle restart, including the loading of a new, certified copy of the Supervisor software, clearing the internal state of the equipment in order to clean up memory untidiness resulting from the shutdown, verifying correct loading of the Supervisor, validating security controls and security parameters, and certifying the system security status by the System Security Officer.

4.4 File control

File control procedures include those for identifying the cognizant agency of each file, scheduling changes for files, modifying access restrictions of files, giving operators access to demountable files, moving files into and out of the computing area, pre-operator handling of files (including mounting and demounting of tapes and discs), and sanitization of files.

4.5 Control of magnetic tapes and discs

These procedures must account for and control the circulation and storage of tapes and discs; their use, reuse, and sanitization; and their classification markings and entrance to and release from the area.

4.6 Control of paper-based media

Procedures for punchcards, forms, paper-tape, and printouts must cover their accountability, classification marking, storage, and entrance to and release from the area. Additionally, manuals, guides, and various system documents must be covered.

4.7 Personnel control

Personnel control procedures include measures for verifying clearances and special-access authorization for personnel entry to each area of the system, visual surveillance of operating and maintenance areas, and logging and escorting of uncleared visitors. The reporting of suspicious behavior and security infractions is included among the personnel control procedures.

4.8 Terminal control

Various procedures are required with respect to the operation of remote terminals. These include provisions for logging user entry to the terminal area, removal of hardcopy, proper marking of hardcopy not marked by the system, clearing of displays, and securing as required during orderly shutdown.

4.9 Security parameter control

Procedures must be provided for authorizing security parameters to be entered into the system; for verifying correct entry; for changing them on the basis of shift, day of the week, etc.; for receiving and processing requests to modify them; and for actions to be taken in case of a system emergency or an external crisis.

4.10 Software control

These include procedures for rigid control and protection of certified copies of the Supervisor and other software bearing on system security or threat to the system, for loading the Supervisor, for making changes to it, and for verifying the changes.

4.11 Maintenance

All maintenance to be performed on hardware or software must be covered by appropriate procedures, including measures for surveillance of maintenance personnel by properly cleared personnel, for verifying with the System Administrator any adjustments made to the system's configuration, and for manually logging all changes and adjustments made or errors discovered

4.12 Certification

Certification procedures should embrace various personnel responsibilities, tests and inspections to be performed and their conduct, the responsibilities of the System Security Officer, etc.

4.13 User aids

The production, distribution, and document control of manuals, guides, job procedure write-ups, etc., must be covered by appropriate procedures; there must be approved ways of conducting personnel training.

4.14 Change of mode

These procedures include the provision of checklists for actions required in changing mode, removal and storage of paper media and demountable files, physical and electronic surveillance of the machine area, purging of printers by running out the paper, purging of punchcard equipment by running out cards, removal or erasure of Supervisor software from the previous mode and proper verification thereof, loading of the Supervisor for the new mode and proper verification thereof, clearing of all storage devices so that residual information from the previous mode does not carry forward, removal of print ribbons from printers and terminal typewriters for storage or destruction, mounting of files for the new mode, and certification of the security status of the new mode.

4.15 Assurance of security control

Security control assurance includes procedures for reporting anomalous behavior of the system or security infractions; for monitoring security controls, including those on communications; for assuring continuity of security control; for devolution of responsibility in case of personnel nonavailability; and for auditing user and system behavior.

5 Conclusion

In the present paper, we present an evaluation through definitions respecting the way they were defined as a groupof the role of management and policy issues in computer security that proposes Rand Report

R-609 within organization. Important to note that an effective and agreed set of management and administrative controls and procedures governing the information's stream to and from the computer system must be added to overall policy guidance and to technical methods.

REFERENCES

[1] Willis Ware. "Security Controls for Computer Systems: Report of Defense Science Board Task Force on Computer Security." Published for the Office of the Secretary of Defense Edited by Willis H. Ware. R-609-1. Reissued October 1979.

[2] Roberts, Larry. "Program Plan for the ARPANET." 8 February 2007 www.ziplink.net/~lroberts/SIGCOMM99_files/frame.htm.

[3] Schell, Roger R., Downey, Peter J., and Popek, Gerald J. Preliminary Notes on theDesign of Secure Military Computer System. January 1973. File, MCI-73-1, ESD/AFSC, Hanscom AFB, Bedford, MA 01731.

[4] Bisbey, Richard, Jr., and Hollingsworth, Dennis. Protection Analysis: Final Report.May 1978. Final report, ISI/SR-78-13, USC/Information Sciences Institute, Marina DelRey, CA 90291.

[5] Grampp, F. T., and Morris, R. H. "UNIX Operating System Security." AT&T BellLaboratories Technical Journal 63, no. 8 (1984): 1649–1672.

[6] Peter Salus. "Net Insecurity: Then and Now (1969–1998)." Sane '98 Online. 19November 1998. Accessed 26 March 2007 from www.nluug.nl/events/sane98/after-math/salus.html.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Investigating Undergraduate Students' Attitudes Towards English Mobile Learning

A Case Study of Moroccan University Students

[1]O.Omari, [2]S.Bourekkadi, [3]K.Slimani, [4]S.Khoulji, [5]M.L. Kerkeb
[1]Faculty of Arts and Humanities, Department of English, Moulay Ismail University, Meknes, Morocco;
[2,4,5]National School of Applied Sciences, Abdelmalek Essaadi University, Tetuan, Morocco;
[3]Faculty of sciences, Ibn tofail University, Kenitra, Morocco.
otmaneomaari@gmail.com

ABSTRACT

The increase of mobile devices, the availability of several features, and the decrease in terms of cost of smartphones made them useful not only for communication, but also for learning. Similarly, the importance of English in higher education makes it one of the most fundamental foreign languages that students want to learn. Many students look for different ways to learn English. Among these ways is Mobile Learning which becomes nowadays a useful tool for language learning due to the evolution of mobile technology. However, little is known about the attitudes of Moroccan students towards English M-learning. Thus, the purpose of this study is to examine the attitudes of using M-learning by university students to learn English. The study involved 286 students representing different universities, levels, and departments. A questionnaire was distributed and interviews were carried out to answer the research questions. Results revealed that university students have positive attitudes towards English M-learning. The total average mean score obtained for the used scale was M=3.80. Additionally, the qualitative findings confirmed some results of the quantitative data and displayed other important findings related to feelings and obstacles towards English M-learning. The findings are of great importance for teachers, practitioners, administrators, and websites/apps programmers.

Keywords: English; Mobile learning; higher-education; university; Morocco.

1 Introduction

The world nowadays is advancing quickly in terms of technology. Everyday we discover new technologies that help us and facilitate our life and our way of living. Education is among the most important fields that know a great amount of technological discoveries. Those discoveries make learning today easier than before and allow many people to reach so much learning materials that were not available in the past. In addition, the increase of technology products such as mobiles, notebooks, tablets, personal computers, electronic translators and others makes the price of these products go down. This allows people to buy these products according to their needs and budget. For instance, mobiles today are everywhere and the majority of people own a mobile or smartphone. More importantly, telecommunication companies make it easy also for mobile owners to access internet through a 3G/4G network facility. Moreover, in many public places Wi-Fi and even Wi-Max are available for free to everyone. All that makes learning easier and

DOI: 10.14738/tmlai.54.3210
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3210

also can be reached everywhere and at any time. In addition, people use these facilities within their portable devices as a learning tool. Today, there is an emerging topic when we talk about learning and technology, which is Mobile Learning (M-Learning).

Nowadays, m-learning is transforming the way students learn in higher education. Students find several ways to access information and can easily find educational resources that enhance their learning process. Meanwhile, according to [1] m-learning is defined as "any educational provision where the sole or dominant technologies are handheld or palmtop devices."(P. 262). This includes mobiles, smartphones, handheld devices, tablets and iPads. Those devices play a crucial role in our daily life and a fundamental role in education when used properly. In fact, there are many ways to use mobiles as an instrument for learning. Many educational applications are available for free and can be used to fulfill different outcomes. For example, on a mobile's online store we can find many applications that help students learn any subject they want. Others can be used to enhance ones Math, Physics, Programming or language skills such as learning English.

On the other hand, due to the fact that English is an important and international language, students are aware now of its importance and tend to look for different ways to develop their English learning skills. In addition, professors and practitioners in the field of education (especially in higher education) encourage students to learn English as it is a language of science that anyone will need while conducting a research and also because it is very important in one's professional life. However, because English is a foreign language in Morocco and some faculties (faculties of science for example) do not include it in their programs, many students look for other ways to learn this crucial language. Meanwhile, the Moroccan government encourages students of all disciplines to learn English, especially students of scientific disciplines. The Moroccan government also launched a program called "Injaz" to support educational development programs. This program provided laptops and tablets with network to students in higher education with low prices. This also allowed students to have their own tablets which are included within the devices that support m-learning. Moreover, this initiative helped students from different social classes to own new technology and use it as a tool of learning. However, this program did not concern all students. Only engineering, master, and PhD students were eligible to benefit from this program once.

In fact, this initiative was made to encourage students to use technology in their learning process; however, there are not enough studies that investigate students' attitudes towards Mobile Learning within the Moroccan context. For that reason and because of the importance of English nowadays, the purpose of this study was to examine students' attitudes towards English Mobile Learning. Thus, the aim of this study was to answer the following 5 research questions:

1. What are students Attitudes towards learning English via Mobile Learning?
2. Are there any differences in students' attitudes in terms of gender, age, major and level of education?
3. Are there any differences in students' attitudes towards English M-learning with regard to their mobile OS and type of Network access?
4. Is there any significant difference among students' attitudes towards English M-Learning in terms of things they use most on their mobiles?

5. What are students' feelings and perceptions towards the use of English m-learning and obstacles that may prevent them from frequently using this technology?

2 Mobile Learning

M-Learning is one of the new and emerging ways of learning using technology. A lot has been said about this topic and students are aware now about it. Some researchers said that m-learning is a simplified way of learning using mobile devices [2, 3]. Additionally, M-learning was defined by [4]as "any sort of learning that happens when the learner is not in a fixed, predetermined location, or learning that happens when the learner takes advantage of the learning opportunities offered by mobile technologies" (p.8). This is to say that M-learning is a way of learning that happens via the use of any kind of mobile devices such as cellphones, tablets, smartphones, PDAs and handheld devices. Meanwhile, the core of the concept is that users can access the information anytime, anywhere from portable devices that they "regard as friendly and personal" (p. 129) [5].

Besides, it is sometimes hard for some people to differentiate between E-learning and M-learning. In fact, though the two concepts differ on their first letters "E" and "M", they share similarities as for online learning materials, participants (learners), and administrators [6]. However, as stated in [6], E-learning and M-learning differ in terms of mobility. On the other hand, this type of learning using mobiles is divided onto formal and informal. Some studies identify formal M-learning as "where learners are engaging with materials developed by a teacher to be used during a program of instruction in an educational environment, highly structured, institutionally sponsored, and generally recognized in terms of a certificate or a credit upon completion" (p.19)[7]. However, informal M-learning is defined as a learning that is unorganized and unstructured [7]. This means that formal M-learning is used within a professional context to serve several academic outcomes or as a requirement of a course; whereas informal M-learning occurs when learners learn using their mobiles to serve their own desires or motivation to learn something not related to their course requirement. However, it is worth mentioning that there is no decisive definition of informal M-learning and that learners can use informal M-learning to serve an academic formal learning outcome [7]. This is to say that M-learning could be used within formal and informal contexts and can also support lifelong learning [8].

Meanwhile, language learning applications are among the most used and known types in M-learning technology [4]. Several studies [9-11] demonstrated the usefulness of M-learning in language learning because it gives students opportunities and it permits self-learning at any time and any place. However, different factors can affect language learning process. For instance, some of the factors that [12] said to influence language learning include: age, intelligence, anxiety, motivation and attitudes. The latter is known to be one of the most affective variables in language learning as demonstrated in a study by [13]. The study concluded that teachers should develop efficient learning strategies to develop students' attitude towards English.

Another study by [14], which explored attitudes and self-efficacy of college students towards M-learning, revealed that attitude is an important variable in the success of mobile learning. The study indicated also that most investigated students had positive attitudes towards M-learning and that the latter increased their motivation to learn English. Meanwhile, a research by [15] was conducted to explore the attitudes of students and educators towards M-learning technology in 5 different universities from the Arab Gulf region (Oman and UAE). The study demonstrated that most students had positive attitudes towards M-

learning with no significant differences with regard to gender, major and level of study. However, significant differences were found in terms of students' age, mobile ownership, and country. In terms of country for instance, learners from UAE were favored. Findings of this study indicated also that the integration of M-learning in higher education could be beneficial to the learning pedagogy within the Arab Gulf countries. Additionally, it was confirmed in a study conducted in Nigeria, which focused on finding out the experiences of students' M-learning in higher education, that mobile features inspire and motivate students to learn more [16].

3 Research Method and Design

The aim of this study was to investigate students' attitudes towards English Mobile learning. Therefore, a mixed research methods using quantitative and qualitative data were employed. We believe that this method will allow us through quantitative data to explore students' attitudes towards learning English using M-learning technology. It will also allow us to discover students' experiences of using M-learning to learn English through interviewing individuals. This research was conducted in February, during the beginning of the second semester of the academic year 2016/2017.

3.1 Research context

The current study was carried out in different Moroccan public universities. Focus was more on students of Sidi Mohamed Ben Abdellah University in Fez and Moulay Ismail University in Meknes. Meanwhile, a copy of the questionnaire was made available online for students from other universities in Morocco. Students from different universities participated in this study, namely from: University Mohamed V (Rabat), Cadi Ayad (Marrakesh), AbdelmalekEssaadi (Tetouan and Martil), and IbnTofail (Kenitra).

3.2 Population and Sampling

A total of 286 students from different universities in Morocco participated in this study. Quantitative data were collected using questionnaires which were completed anonymously. The qualitative data were used to answer the first 4 research questions. On the other hand, interviews were used for qualitative data to answer the last research question (question number 5). Questionnaires were administered during February 2017 in two universities (Sidi Mohamed Ben Abdellah and Moulay Ismail University). Sixty-five questionnaires were distributed randomly at Sidi Mohamed Ben Abdellah University and forty at Moulay Ismail University. In addition, an online version of the questionnaire using "Google forms" was created and shared via Social Media to several groups in order to receive responses of students from other universities in Morocco.

Meanwhile, a demographic information form was used to collect students' information such as gender, age, institution, level of education, department, and city. Other variables were added to the demographic information form in order to find out to what extent they may affect students' attitudes toward learning English via M-learning. These variables include questions about smartphone ownership, access to internet, type of operating system (OS), and the most three things students frequently use on their mobiles. Besides, only undergraduate students were asked to fill out the questionnaire voluntarily, which took them less than 5 minutes. They were assured that their information will be kept strictly confidential.

3.3 Study instrument

In order to investigate university students' attitudes towards learning English using mobile learning technology, a survey was adapted from [15]. The authors developed an instrument of 28 items by adopting several items from relevant studies. The instrument was later on reduced to 10 items, tested, and validated by the authors. Due to the specificity of the current study, some additional modifications were carried out. For instance, the term 'Mobile Learning' was used in the current study instead of 'Mobile Technology'. Additionally, the items were modified to fit the English learning context. The word English was added to all items to collect participants' responses concerning their attitudes towards learning English via M-learning. For example, the first item "Mobile technology is a useful tool for my study" was modified to "Mobile Learning is a useful tool for me to study English".

Before the part of the scale, participants were introduced to the meaning of Mobile Learning in order to insure their understanding of the concept. After that, participants were asked to rate how much they agree or disagree with each statement. A five-point-Likert Scale ranging from (1) Strongly Disagree to (5) Strongly Agree was used to collect students' attitudes towards English M-learning. To get a better understanding of students' attitudes towards the use of M-learning to learn English, we interviewed 21 volunteer students (13 females and 8 males). The interviews were conducted with random selection of students. Semi-structured questions were used in the interviews which lasted about 10 minutes each. Besides, due to the modifications made in the scale section, we tested the reliability of the scale using SPSS to determine its validity. Alpha method was used and a coefficient of 0.84 was obtained.

3.4 Demographic profile

The demographic results indicate that 47.6% of the participants were males and 52.4% were females. Concerning age, the largest age group falls within 16 – 20 years represented by 59.8%, participants aged between 21 – 25 years represented by 34.6%, whereas the smallest age group was that of participants aged between 26 – 30 years represented by 5.6%. On the other hand, Sidi Mohamed Ben Abdellah University in Fez was represented by the largest number of participants (89, 31.1%). The other institutions were represented as follows: Moulay Ismail University (55, 19.2%), Mohamed V University (31, 10.8%), Abdelmalek Essaadi Univeristy in Tetuan (ENSA, Science)(30, 10.5%), Abdelmalek Essaadi Univeristy in Martil (29, 10.1%), Ibn Tofail University (28, 9.8%), and Cadi Ayad (24, 8.4%).

Participants represented different levels and were from different departments. Most students were second and first year undergraduate students (respectively 42% and 37.8%). In terms of departments, students from the department of English represented more than half of the sampling (52.1%). Other departments were as follows: Science (9.4%), Economics (24%), Physics (10.1%), French (7.3%), Math (4.2%) and from other departments (8.4%).

Meanwhile, in response to smartphone ownership, 95% of participants declared that they have a smartphone. However, only 5% said that they do not own a smartphone. In terms of the mobile's operating system, 64% said they have an android, 23% iOS, 8% windows, and 5% said that they have another operating system different from the previously mentioned. Concerning internet access, approximately 38% of participants said that they access internet through 3G/4G only, 19% through Wi-Fi only, 32% use both 3G/4G and Wi-Fi, and about 10% said that they do not have access to internet.

4 Results

To make the analysis of the findings easy and simple, we will divide the results into two parts. The first part will focus on the findings of the quantitative data (survey), and the second part will display the findings of the qualitative data (interviews).

4.1 Quantitative data

4.1.1 What are students Attitudes towards learning English via Mobile Learning?

This research question depends on participants' overall ratings on their attitudes towards English M-learning scale. The overall mean of the scale was 3.80 with a standard deviation of 0.476. Students' highest overall score was that of "Mobile Learning is a useful tool for me to study English" (M = 4.14). Table I reveals the mean and standard deviation of each item. As displayed, the other items range from 3.15 to 4.

Table 1. Students' attitudes to English m-learning scale

Items	Mean	Std. Deviation
1. Mobile Learning is a useful tool for me to study English.	4.14	0.603
2. Mobile Learning can offer me opportunities for English communication and team-working.	3.97	0.505
3. Mobile Learning can help me find resources related to English learning.	3.87	0.556
4. Mobile Learning can bring many opportunities to my English learning process.	3.73	0.573
5. Mobile Learning can help me access English course-material anytime anywhere.	3.77	0.591
6. Mobile Learning can be an easy way to get feedback and notifications from instructors.	3.15	0.551
7. Mobile Learning can help me exchange the English course-material with my friends	3.67	0.617
8. Mobile Apps can help me manage my English Learning.	3.77	0.633
9. Mobile Learning can help me to do my English coursework.	3.86	0.637
10. Mobile Learning can help me develop my English learning skills.	4.07	0.590

4.1.1 Are there any differences in students' attitudes in terms of gender, age, major and level of education?

After calculating the means of students' attitudes to learn English via M-learning with regard to gender, we found that there are no differences in students' attitudes in terms of gender. The score of males was M= 3.79 and the score of females was M= 3.82. Similar to that, no big differences were found in terms of age. Both students who are aged between 16 and 20 scored a mean of M= 3.82, and those aged between 21 and 25 scored a mean of M= 3.78. Meanwhile, first year students (M= 3.89) showed a slight difference compared with second year and third year undergraduate students (Respectively M= 3.76, M= 3.72). Similarly, students fromdepartments of English, French and Physics demonstrated a small difference compared with students representing other departments as displayed in Table III.

4.1.2 Are there any differences in students' attitudes towards English M-learning with regard to their mobile OS and type of Network access?

Again, results demonstrated that there are no big differences in students' attitudes towards learning English via M-learning with regard to the used operating system. The slight difference as demonstrated in

Table IV is within Android and iOS compared with the other operating systems. On the other hand, students who declared that they access internet using both 3G/4G network and Wi-Fi scored (M= 4.05) more than those who access internet through 3G/4G only, Wi-Fi only or those who do not have access to internet as demonstrated in Table V.

4.1.3 Are there any significant differences among students' attitudes towards learning English via M-Learning in terms of things they use most on their mobiles?

To answer this question, students were asked to select the most three things they frequently use on their mobiles. As demonstrated in Table VI, the majority of students declared that the three things they use most on their mobiles are social networking, instant messaging, and listening to music. In terms of their attitudes to learn English via M-learning and the things they use most on their mobiles, we notice that students who declared that accessing email is one of the most three things they use scored a higher level of attitude (M= 4.14). Meanwhile, the other items were about similar ranging from a score of M= 3.61 to 3.87.

Table 2. Means according to major

Department/Major	Mean	N	Std. Deviation
English	3.98	149	0. 466
Physics	3.95	29	0.478
French	3.91	21	0.382
Economics	3.73	24	0.410
Science	3.66	27	0.524
Maths	3.56	12	0.536
Other	3.74	24	0.417
Total	3.80	286	0.476

Table 3. Means according to type of OS

Type of mobile OS	Mean	N	Std. Deviation	% of Total N
Android	3.96	184	0.472	64.3%
Ios	3.92	64	0.494	22.4%
Windows	3.70	23	0.464	8%
Other	3.57	15	0.470	5.2%

Table 4. Means according to internet access

Internet Access	Mean	N	Std. Deviation	% of Total N
3G/4G	3.72	109	0.451	38.1%
Wifi	3.78	55	0.551	19.2%
Both	4.05	92	0.488	32.1%
No internet	3.61	30	0.455	10.5%

Table 5. Things students use most on their mobiles and means of their attitudes towards Englishm-learning

Most used things[a]	Responses		Percent of Cases	Attitude Mean
	N	Percent		
Social Networking	261	30.9%	91.3%	3.81
Instant messaging	236	28.0%	82.5%	3.82
Listening to music	121	14.4%	42.3%	3.65
Searching for information	91	10.8%	31.8%	3.87
Accessing Email	54	6.4%	18.9%	4.14
Watching videos/movies	20	2.4%	7.0%	3.64
Playing games	20	2.4%	7.0%	3.61
Reading (articles, books…)	13	1.5%	4.6%	3.80
Other	27	3.2%	9.4%	3.61
Total	843	100.0%	294.8%	

a. Dichotomy group tabulated at value 1

4.2 Qualitative data

As mentioned earlier in the methodology section, 21 volunteer students accepted to participate in the interview. A descriptive analysis is carried out to identify students' attitudes and perceptions concerning learning English using M-learning technology.

Q1: What is your opinion concerning the importance of learning English?

The first question was about the importance of English in relation to students' field of study and future career. For this question, all participants confirmed that English for them is very important and that they want to enhance their English learning skills. Participants from the department of English said that English is their major, so they are required to look for ways to develop their English skills. On the other hand, participants representing other departments declared that English for them is crucial because they know that they will need it either during their graduate studies or when they want to get a job. Some participants said, "English for me is very important because it will help me find a suitable career after I graduate", "English is very important because it is the language of science, learning this language will help me when I want to conduct my research paper", "Except from the fact that it is a universal language, I have always loved it since I was in high school".

4.2.1 Q2: What do you think about using your mobile to learn English?

The second question was about students' perceptions on the use of M-learning technology to learn English. Most participants (17) indicated that this technology is very useful for them and that it can help them learn the language whenever they want and at their ease. Other students said that M-learning to learn English "saves time and money" and "it is a lot of fun". Meanwhile, some students suggested that M-learning would be more interesting and beneficial if teachers afford them with helpful electronic materials. On the other hand, others declared that learning via mobile disturbs them, especially when they are using internet. Table VII demonstrates different opinions of students concerning that question.

Table 6. Students' statements for interview Q2

Statements	Frequency (N=21)
Using technology to learn English is very helpful to me.	17
When I have time, I go online and look for something that will help me learn English.	15
Learning English via mobile is a lot of fun.	13
Learning English using my mobile will save me a lot of money.	13
With mobile, I can get a lot of information quickly and at any time.	11
Learning via mobile will be more beneficial if teachers can afford us with useful materials that will help us.	9
I get distracted when I use my mobile to learn.	6
I find what I study in class enough. I use my mobile just to check for definitions.	5
I don't think this is a good idea.	3

4.2.2 How do you feel about using your mobile to learn English or to support your English learning course?

Students in this question were asked to describe their feelings concerning learning English using their mobiles. The majority of participants (14) stated that they have already the previous experience using their mobiles to support their English language learning skills. They also said that they enjoyed a lot learning using their mobiles because it is easy and quick, and also because they do not need to wait for their computers to turn on. Some (12) also said that the use of M-learning to learn English helps them and will help them to develop their language skills especially in terms of listening, reading and vocabulary skills. However, some students (6) declared that they feel anxious because they cannot find useful materials that will enhance their English learning skills.

4.2.3 Is there anything that makes you sometimes unwilling to use your mobile to learn English?

To identify obstacles that may prevent learners from using their mobiles to learn English, participants were asked to state anything that makes them hesitating to use their mobiles. As displayed in Table VIII, one of the most reasons that make students do not frequently use or hesitate to use their mobiles to learn English is 'bad network'. Another reason was related to offline mode. Some students said that they cannot have internet network every day, which prevent them from accessing the content frequently. They also added that though there are some applications providing offline mode, they are still limited in terms of content. Meanwhile, some participants declared that most websites and applications do not provide a placement test and this makes it difficult for them to find the appropriate material for their levels.

Meanwhile, some students stated that a lot of advertisements on mobile applications or websites prevent them from trying to learn English via M-learning. One of the students said "When I visit a website and I find a lot of advertisements, I just quit". Another said "Sometimes I visit a website, and when I want to change the task it takes me to another website for advertisement reason".

Table 7. Students' statements for interview Q4

Statements and reasons	Frequency (N=21)
Bad internet network.	18
Most applications/websites I used donot provide a placement test.	14
Lack of user feedback feature	12
Bad design and interface.	12
Apps do not support offline mode.	11
Bad app performance (Slow response, long load times, crashes).	10
A lot of advertisements.	10
Apps requesting access to private content.	8
I don't know a useful website or app that can help me learn English.	7
Some apps need big space.	6
I find it difficult to download or copy/paste content from my mobile that I like for future use.	5
When I use my mobile, I can't focus on what I am learning especially when someone sends me a message on WhatsApp or Facebook.	5
Sometimes it is difficult to follow the instructions or there are no instructions.	4

5 Discussion

The goal of this study was to examine undergraduate students' attitudes towards using Mobile Learning to learn English. A survey was distributed to different students representing different universities, gender,

ages, levels, and departments. Additionally, some interviews were carried out to better understand students' attitudes and the barriers that prevent them from using Mobile technology to learn English or enhance their English skills.

In general, the overall average score of students' attitudes towards learning English via M-learning (M=3.80) shows that students have positive attitudes. This is to say that students agree that M-learning is a useful tool to learn English. These results were in conformity with other findings of other studies conducted in different countries revealing the positive attitudes of students towards Mobile Learning [14, 17-19]. In terms of items, we notice that the first item about the usefulness of M-learning to learn English was the highest (M= 4.14) followed by item number 10 about the helpfulness of M-learning in the development of the English learning process (M= 4.07). It is clear here that both items have got the highest scores. This means that students see M-learning as a tool that can help them enhance and develop their English learning skills. On the contrary, the 6th item about supporting communication with instructors was the lowest with regard to the overall average scores (M= 3.15). When we compare this item with items number 1 and 2, we notice that there is a big difference between them. For instance, the 2nd item which is also about the opportunities that M-learning provide to communicate with colleagues (M= 3.97) has a higher score than the 6th item which falls in the same category of communication. Thus, we can conclude that there maybe a problem concerning students' attitudes towards communication with instructors using M-learning.

Concerning the differences of students' attitudes to learn English using mobile learning in terms of gender and age, the results confirmed that there are no significant differences. This means that students' attitudes are equal and that gender or age does not affect their attitude to learn English via M-learning. On the other hand, a slight difference was found between students' attitudes overall score in terms of level of education. First year students' score was higher than second and third year; however, this cannot be considered as a significant difference. Meanwhile, in terms of major, the results demonstrated that students from the departments of English, physics, and French scored higher than students from the other departments. In fact, it sounds reasonable for students from the department of English to have a higher score because English is their major. However, we can say that this is a positive result because there isn't much research that tackles attitudes of students from the department of English in Morocco. Meanwhile, it is very important to highlight the score of students from the department of physics (M= 3.95) concerning their attitudes towards M-learning to learn English, especially because we know that English is not included in undergraduate programs at faculties of science such the case in Fez. Hence, we suggest that this result should be taken into consideration for practitioners and people in power to include English in all undergraduate levels and in all faculties without any exception.

Besides, the results revealed that most students own a smartphone and can access internet. This can be considered as a good result, especially because we are in an age of technology. Similarly, most students were found to own a smartphone with an Android or iOS operating system. Meanwhile, the results displayed that there is a difference between students' attitudes towards learning English using M-learning in terms of the used operating system. Students who use Android or iOS displayed higher attitudes than the others. Similarly, students who access internet via both Wi-Fi and 3G/4G displayed also a high level of attitude compared with those who access internet via 3G/4G only, Wi-Fi only, or those who don't have

access to internet. These results show that the way used to access internet and the operating system can affect students' attitudes towards learning English using M-learning. Thus, we can say that students who always have access to internet (via Wi-Fi and 3G/4G) may experience higher attitudes towards learning English via M-learning. Similarly, students who own an Android or iOS operating system may have higher attitudes than others using other operating systems. This is maybe due to the fact of those operating systems features and their popularity in Morocco and worldwide. This can be also due to the ease of use and the richness of the available applications in the play stores of those two operating systems.

On the other hand, results for the fourth research question demonstrated that the most frequently used things that a student use are social networking 31%, instant messaging 28% and listening to music 14%. However, we notice that in terms of students' attitudes towards English M-learning, students who declared that accessing their emails was among the most things they frequently use scored higher than the others. Thus, we can say that students who access their emails frequently may have more attitudes to learn English via M-learning. Additionally, the use of emails is considered as part of the learning process when it is linked with formal learning. Many studies revealed the effectiveness of emails use to facilitate class activities [20-22] and collaborative work [23]. Moreover, other studies [24, 25] confirmed that email use enhances students' attitudes and interest towards the learning process.

Meanwhile, more focus should be on enhancing students M-learning attitudes towards communication with instructors. More research should be carried out to examine teachers' attitudes towards the use of technology, especially because evidence [26, 27] revealed that teachers' positive attitudes towards the use of technology help them integrate technology to support the learning process. That way, communication between students and instructors via technology will be enhanced.

With regard to qualitative findings, we can say that the results provided significant and useful information that could help understand students' attitudes and perceptions towards using M-learning to learn English. Meanwhile, it is very essential to highlight students' awareness about the importance of English nowadays. All students provided positive comments and ensured that learning English is crucial. On the other hand, the obtained results for the second question displayed that the majority of students consider M-learning a useful tool to learn English. This finding confirms the results obtained from the scale in this study. However, we should not forget that other students need the support of their teachers to help them find useful materials that could be beneficial to learn English through the mobile. Thus, teachers also should be involved in this matter. Additionally, some students indicated that they are easily distracted when they use their mobiles as a tool of learning. This may be because they do not link learning activities with the use of mobile devices [28]. Also, some students may be easily disturbed because learning the language is not a priority for them or because they think of the mobile as a tool of entertainment only. As concluded by [29] (in their study investigating the experience of students while using portable devices to learn) they said that students need help to recognize the value of portable devices in learning, and particularly mobile learning.

Findings of the third question of the qualitative data confirmed again the findings of the used scale and also the previous question. Most interviewed students displayed positive feelings about using the mobile to learn English. They also confirmed that mobile learning is an effective way of learning English which is quick, easy, and enjoyable. On the other hand, results obtained from the last question of the qualitative data displayed some important barriers that disturb and prevent students from using M-learning frequently. As mentioned in the result section, the interviewed students declared that bad network is

among the most obstacles that prevent them from using M-learning frequently. That is why some of them added that apps lacking the offline mode are sometimes useless. This is due to the fact that most mobile apps that support the offline mode provide limited content. Meanwhile, the findings displayed the importance of providing a feedback feature and a placement test for language learners. In addition, an important number of participants declared that the design and interfaces may stand as obstacles for them from using M-learning. In fact, these findings would be of great importance for mobile apps and websites designers. Thus, it is important to take these obstacles into consideration while developing a website or an application for language learning purposes.

6 Conclusion

In general, this study examined the attitudes and perceptions of undergraduate students towards learning English using M-learning. The study was conducted in several universities in Morocco and participants represented different departments and cities. Quantitative and qualitative data were carried out to understand how students perceive the use of their mobiles to learn English. The analysis of students' attitudes indicated that they have positive attitudes towards the use of M-learning as a tool of learning English. No big differences were found for students' attitudes in terms of gender, age, and level of education. Only a tenuous difference was found within students' major, operating system and network used. Meanwhile, the study revealed that students who check their emails frequently via mobile were found to have higher attitudes towards learning English via M-learning than the others. Additionally, the analysis of the findings displayed that students are aware of the importance of English and that they have positive feelings towards the use of M-learning to enhance their English language learning. Meanwhile, this study provided some of the obstacles that may prevent or make students hesitate to use M-learning. These findings will enrich the available literature and may be of great benefit for practitioners as well as websites and mobile applications programmers in the future.

The limitation of this study is that it was conducted in English. The survey was administered in English and also the interviews were carried out using English, except for some cases when some explanations were needed to be done using the mother tongue (Moroccan dialect). Another limitation is that this study was not conducted in all Moroccan public universities, so the results should not be generalized. Meanwhile, the findings of this study could have important implications for teachers, institutions, programmers, and learners who would like to learn English through their mobiles. The next step should be the integration of mobile learning in university classrooms using some local applications that are supported by the Moroccan institutions and universities. Future research should focus on identifying students' interest and preparedness to participate in English M-learning. Moreover, correlations should be carried out between students' attitudes towards English M-learning and the things they use most on their mobiles. In addition, more research is needed about teachers' attitudes and perceptions towards using and integrating M-learning in their English teaching courses.

ACKNOWLEDGMENT

We are very grateful to all students who agreed to participate in this study. Their willingness to give their time so generously has been very much appreciated. We would like also to express our sincere thanks to CNRST Morocco.

REFERENCES

[1] J.Traxler, Defining mobile learning. in IADIS International Conference Mobile Learning. Qawra, Malta, pp 261-266, 2005.

[2] A. Herrington, and J. Herrington, Authentic mobile learning in higher education, in AARE 2007 International Educational Research Conference. Association for Research in Education: Fremantle, Western Australia, 2007.

[3] J.-H.Valk, A.T. Rashid, and L. Elder, Using mobile phones to improve educational outcomes: An analysis of evidence from Asia.11(1): p. 24, 2010.

[4] C. O'Malley, et al., Guidelines for learning/teaching/tutoring in a mobile environment, 2005.

[5] J. Traxler, Defining, Discussing and Evaluating Mobile Learning: The moving finger writes and having writ.The International Review of Research in Open and Distributed Learning, 8(2) 2007.

[6] D. Najima, and A. Rachida, An Adaptation of E-learning Standards to M-learning. International Journal of Interactive Mobile Technologies. 2(3): p. 11-16, 2008.

[7] J.Gikas, and M.M. Grant, Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. The Internet and Higher Education. 19: p. 18-26,2013.

[8] N.L.Beddall-Hill, and J. Raper, Mobile devices as 'boundary objects' on field trips. Journal of the Research Center for Educational Technology. 6(1): p. 28-46,2010.

[9] C. Evans, The effectiveness of m-learning in the form of podcast revision lectures in higher education. Computers & Education. 50(2): p. 491-498,2008.

[10] M. Lu, Effectiveness of vocabulary learning via mobile phone. Journal of Computer Assisted Learning. 24(6): p. 515-525,2008.

[11] P. Thornton, and C. Houser, Using mobile phones in English education in Japan. Journal of Computer Assisted Learning. 21(3): p. 217-228, 2005.

[12] M. Al-Emran, H.M. Elsherif, and K. Shaalan, Investigating attitudes towards the use of mobile learning in higher education. Computers in Human Behavior. 56: p. 93-102,2016.

[13] S.S. Oyelere, J. Suhonen, S.A. Shonola, &M.S. Joy, Discovering students mobile learning experiences in higher education in Nigeria. in 2016 IEEE Frontiers in Education Conference (FIE). 2016.

[14] F.N. Al-Fahad, Students' attitudes and perceptions towards the effectiveness of mobile learning in King Saud University, Saudi Arabia. TOJET: The Turkish Online Journal of Educational Technology. 8(2), 2009.

[15] N.Cavus, and D. Ibrahim, m-Learning: An experiment in using SMS to support learning new English language words. British Journal of Educational Technology. 40(1): p. 78-91,2009.

[16] R.M. Jaradat, Students' Attitudes and Perceptions towards using m-learning for French Language Learning: A case study on Princess Nora University. Int. J. Learn. Man. Sys. 2(1): p. 33-44,2014.

[17] L.E.S.de Montes, and C.L. Gonzales, Been There, Done That: Reaching Teachers Through Distance Education [*]. Journal of Technology and Teacher Education. 8(4): p. 351-351,2000.

[18] G. Lawrence, The Use of E-mail as a Tool to Enhance Second Language Education Programs: An Example from a Core French Classroom. Canadian Modern Language Review. 58(3): p. 465-472,2002.

[19] D.M. Poole, An email activity: preservice teachers' perceptions of authenticity. Journal of Technology and Teacher Education. 8(1): p. 13-28,2000.

[20] J. Grünberg, and A. Armellini, Teacher collegiality and electronic communication: A study of the collaborative uses of email by secondary school teachers in Uruguay. British Journal of Educational Technology. 35(5): p. 597-606,2004.

[21] T.L.Clingerman, and J.M. Bernard, An Investigation of the Use of E-Mail as a Supplemental Modality for Clinical Supervision. Counselor Education and Supervision. 44(2): p. 82-95,2004.

[22] P. Boxie, Making a cyber literacy connection from the storage room to the college room. Reading Horizons. 45(2): p. 127,2004

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Comparative Analysis of Three Metaheuristics For Solving The Travelling Salesman Problem

Safaa Bouzidi, Moahmmed Essaid Riffi, Abdelhamid Bouzidi
Dept. of Computer Science, Faculty of Science, ChouaibDouakkali University
El Jadida, Morocco
sfbouzidi@gmail.com, said@riffi.fr, mr.abdelhamid.bouzidi@gmail.com

This research paper aims to do a comparative study of three recent optimization metaheuristic approaches that had been applied to solve the NP-hard optimization problem called the travelling salesman problem. The three recent metaheuristics in study are cuckoo search algorithm, cat swarm optimization algorithm and bat-inspired algorithm. To compare the performances of these methods, the three metaheuristics are applied to solve some benchmark instances of TSPLIB. The obtained results are collected and the error percentage is calculated. The discussion, will present which method is more efficient to solve the real application based on the travelling salesman problem.

Keywords- Metaheuristic; Cat Swarm Optimization; Bat-inspired algorithm; Cuckoo Search; Travelling salesman problem; NP-hard

1 Introduction

The Travelling salesman problem (TSP) is one of the best known in the operations research, studied since the 19th century, and was introduced at first as a game by Wiliam Rowan Hamilton and Thomas Penyngton Kirkman [1]. This problem aim is to find the minimal possible path, by starting from a city, and visiting atonce each city in the problem, and return to its city of start. In general, we try to minimize the total travel time and total distance. In addition, this problem is NP-hard problem [2].

The importance to the resolution of traveling salesman problem appears in various complex application area in our real life, such as:

- Drilling of printed circuit boards
- Overhauling gas turbine engines
- Computer wiring
- Vehicle routing

In front of its importance, and the difficulty to solve it. The researchers had try to solve it by numerous methods. One of them is the computational intelligence based on nature observation and study, called the metaheuristics. There is also the exact methods, but it can only solve the TSP instances problem with minimal size, but when the number of city is high, its impossible to find the global solution. But after using the metaheuristic method, it's possible to find the best optimal solution in a reasonable execution time, as the simulated annealing [3], Tabu search [4], Harmony search [5], cuckoo search algorithm [6], Genetic algorithm [7], Ant colony optimization [8], Particle swarm optimization [9], Bee colony optimization [10], cat swarm optimization algorithm [11], bat swarm optimization [12]. To improve the efficiency of some

DOI: 10.14738/tmlai.54.3211
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3211

methods, the researchers had proposed some hybrid methods that had proven their efficiency to solve the TSP, such as, the hybrid genetic algorithm [13], hybrid ant-colony optimization [14], hybrid particle swarm optimization [15], hybrid bee colony optimization [16]. And more. However, the problem is to know which one is more efficient so it can be applied for solving a real TSP problem based.

The aims of this paper is a comparison study, between three recent algorithms to solve the TSP, and know which one is efficient to be applied in others, in order to apply them to a real application based on the travelling salesman problem. The rest of this research paper is organized as follows: section II presents the formulation of TSP. Section III, is a brief description of the three metaheuristics in study. Section IV shows the result of this study, and a discussion. Finally, the conclusion.

2 Traveling Salesman Problem

Given a graph G = (X, E), where X = {1,...,n} is a set of nodes, each node presents a city, and E = {1,...,m} presents a set of edges, each edge presents a path between two selected cities. and costs, dij , presented the arcs distance between vertices i and j, the TSP consists in finding the Hamiltonian cycle of G with the best minimal total length D.

$$D= \sum_{i=1}^{n} \sum_{i=1}^{n} d_{ij} \quad (1)$$

3 Metaheuristics Description

This section is dedicated to describe the three metaheuristic that have been recently improved to be able to solve the NP-combinatorial optimization problem, in specal the TSP.

These methods are :

- The cat swarm optimization algorithm.
- The bat swarm optimization algorithm.
- The Cukoo search algorithm.

3.1 Cat swarm optimization

The metaheuristic Cat swarm optimization (CSO) is an evolutionary method proposed in 2006 by Chu and Tsai [17]. To describe the natural behaviors of cats. The real cat spends the majority of his life, about Seventeen five percent; at rest, attentive and vigilant in observing its surroundings for his next move. Moreover, twenty percent of its life in chase, by moving quickly to chase a pray or any moving object.

In the CSO algorithm, each cat can be considered as a candidate solution. There are two mode; the seeking mode (SM) and the tracing mode (TM). The seeking mode presents the real cat at rest. The tracing mode presents the cat when it traces its path, according to its own velocity to chasing a prey or any moving object. To combine these two modes, the mixture ratio (MR) is defined.

The characteristic of each cat are:

- The position presents the solution.
- the velocity to update the position.
- the flag defines the mode that characterizes a cat.

The description of the total algorithm process of CSO is:

BEGIN
Create N cats;
Initialize the position, velocity, and flag of every cat;
Repeat
Evaluate the position of each cat;
keep the position of the cat with the best fitness;
value into Gbest parameter;
For(each cat_k in the swarm)
If (flag of cat_k is SM)
Begin
Make j copies of cat $_{k;}$
If (SPC==true)
Begin
includes cat $_k$ as a candidate;
j=SMP-1;
End
ELSE
j= SMP;
EndIf
For(each copy)
Select a number of dimensions based on CDC;
Update their value using SRD percent of their
current value;
Evaluate the fitness of each copy;
For(each cat)

$$P_k = \frac{|FS_k - FS_{min}|}{FS_{max} - FS_{min}}$$

Randomly select a new position for cat $_k$;
EndFor
END
ELSE
Begin
$V_i = V_i + r * c * (X_{best} - X_i)$ (2)
$X_i = X_i + V_i$(3)
End
EndIf
Reinitialize the flag of each cat;
Until (Condition is satisfied)
END

Where, in equation (2):

- **X** $_i$The position of the selected cat_i.
- $\mathbf{X}_{best}$: is the best solution/position of the cat who has the best fitness value.
- **V** $_i$: The old speed value (current value).
- **c**: is a constant.
- **r**: a random value in the range [0, 1]

Moreover, in equation (3):

- **X** $_k$: The position of selected cat $_i$
- **V** $_k$: The velocity of cat $_i$

The adaptation of operation used in CSO algorithm:

- Movement: the movement of cat $_k$ is a swap applied to solution of this cat.
- Addition:
- The addition between position **X** and a velocity **V**, is applying the swap in velocity **V** to position **X**, the result is a new position **x'**.
- The addition between two velocities **v** and **v'**, is a new velocity containing all the couple of swaps of **v** and **v'**.
- Subtraction: the result of the subtraction between two positions **x** and **x'** is a velocity **v**, it is the opposite of addition:

$$\mathbf{x} + \mathbf{v} = \mathbf{x'} \Leftrightarrow \mathbf{x'} - \mathbf{x} = \mathbf{v}$$

- Multiplication: the multiplication is performed between a float value and velocity, the result is a velocity. The different possible cases according to the real **k** are:

- If k = 0: k * v = 0
- If (k>0 & k<=1) :r * v = $(i_k,j_k)[k : 0 \rightarrow (c*|v|)]$
- If k>1: then we separate. Decimal and integer part, k = n +x. Where n is the integer part of r, and x corresponds to the decimal parts. We will then return each party to the previous cases.
- If k<0: k * v= (-k)*¬v. Now (-k) >0, and you will consider one of the previous cases.

This methode was improved in 2012, by A.bouzidi and M.E. RIFFI [11], to solve the travelling salesman problem.

3.2 Bat-inspired algorithm

Bat algorithm (BA) is a bio-inspired algorithm and stochastic optimization technique developed by Yang in 2010 [18]. This metaheuristic is inspired behavior of bats. In nature, bats emit ultrasound pulses to the surrounding environment with hunting and navigation purposes. After the emission of these pulses, bats hear the echoes, and based on them they can locate themselves and identify obstacles and preys. Furthermore, each bat of the swarm is able to find the most nutritious areas performing an individual search, or moving towards a nutritious location previously found by the swarm.

The basic Bat algorithm developed by Xin-She Yang [18] is described in following steps:

```
BEGIN
    Define the objective function f(x);
    Initialize the bat population X = (x1,x2,...,xn);
    for each bat xi in the population do
      Initialize the pulse rate ri, velocity vi and loudness Ai;
      Define the pulse frequency fi at xi;
    End
    Repeat
    for each bat xi in the population do
       Generate new solutions through Equations 4 and 5;
    if rand>rithen
         Select one solution among the best ones;
         Generate a local solution around the best one;
    End
    if rand<Ai and f(xi)<f(x∗) then
         Accept the new solution;
```

```
        Increase ri and reduce Ai;
      end
      end
      until termination criterion not reached
      Rank the bats and return the current best bat of the
      population;
    END
```

Y.Saji and ME RIFFI [12] have improved in 2015 the bat-algorithm; by solving the NP-hard combinatorial optimization. That had proven its efficiency to solve the travelling salesman problem.

For each virtual bats, the movement is given by updating their velocities v_i^{t+1} and positions x_i^{t+1} at time step t+1 using (4) and (5), as follows:

$$v_i^{t+1} = \chi(x_i^t, x_*, f_i) \quad (4)$$

$$x_i^{t+1} = \varphi(x_i^t, v_i^{t+1}) \quad (5)$$

The function χ is a crossover function that has three input arguments and returns a set of permutations reached by applying 2-exchange crossover mechanism described in [12].

The function ϕ is a function to sort x_i^t while taking into account the set of permutations of v_i^{t+1}.

3.3 Cuckoo search algorithme

The Cuckoo Search (CS) algorithm is a metaheuristic introduced first in 2009 by Xin-She Yang and Suash Deb [19], bio inspired by the breeding behavior of cuckoos. It was inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds (of other species). Some host birds can engage direct conflict with the intruding cuckoos. Some cuckoo species such as the New World brood-parasitic Tapera have evolved in such a way that female parasitic cuckoos are often very specialized in the mimicry in colors and pattern of the eggs of a few chosen host species.

In the CS method, each egg in a nest represents a solution, and a cuckoo egg represents a new solution. The aim is to use the new and potentially better solutions (cuckoos) to replace a not-so-good solution in the nests. In the simplest form, each nest has one egg. The algorithm can be extended to more complicated cases in which each nest has multiple eggs representing a set of solutions.

The process of the CS is as follow:

```
BEGIN
 Objective function f(x),x=(x1,...,xd)^T
 Generate initial population of n host nests xi (i = 1,2,..., n)
while (t <MaxGeneration) or (stop criterion)
Get a cuckoo randomly by Levy flight;
Evaluate its quality/fitness Fi;
   Choose a nest among n (say, j) randomly;
if (Fi > Fj)
replace j by the new solution;
End if
   A fraction (pa) of worse nestsare abandoned and new
ones are built;
   Keep the best solutions(or nests with quality solutions);
   Rank the solutions and find the current best;
Endwhile
 Postprocess results and visualization;
END
```

A cuckoo i generates new solution$x_i^{(t+1)}$by a Levy flight, according to (6) :

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \oplus Levy(s,\lambda) \quad (6)$$

Where$\alpha$is the step size that follows the Levy distribution in (7):

$$Levy(s,\lambda) \sim s^{-\lambda} \quad (7)$$

Which has an infinite variance with an infinite mean [19]. Here,s is step size drawn from a Levy distribution.

Some rules about CS should be respected, are:

(1) Each cuckoo lays one egg at a time and selects a nest randomly.
(2) The best nest with the highest quality of egg can pass onto the new generations.
(3) The number of host nests is fixed, and the egg laid by a cuckoo can be discovered by the host bird with a probability pa$\in$[0 , 1] .

The improved Cuckoo search algorithm had succeeded to solve the travelling salesman problem by A. Ouaarab and al. [6].

4 Metaheuristique to Solve Tsp

This part is devoted to the collect and analysis of the results after applying the three metaheuristics to solve fourteen-benchmark instance of TSPLIB [20]. The collected result will show which method is more efficient to solve the real application based TSP.

The following table shows the collected result of each instances benchmark taken from TSPLIB, the size of the instance (number of edge in the graph), the best-known optimal global value, and for each method, the best, and the worst result got in ten iteration. In the last, the relative percentage error (RPD) calculated in ten iteration as follow:

Table 9 : results by the application to solve some banchmark instances of TSP

Instance	size	Optimum	CSO algorithm			Bat algorithm			Cuckoo-search		
			BestR	WorstR	RPD(%)	BestR	WorstR	RPD(%)	BestR	WorstR	RPD(%)
eil51	51	426	426	427	0,07	426	426	0.00	426	426	0.00
berlin52	52	7542	7542	7693	1.00	7542	7542	0.00	7542	7542	0.00
st70	70	675	675	682	0.51	675	675	0.00	675	675	0.00
eil76	76	538	538	549	1,02	538	542	0.14	538	539	0.17
kroA100	100	21282	21282	21717	1,0220	21282	21282	0.00	21282	21282	0.00
kroB100	100	22141	22141	23014	1,9715	22141	22141	0.00	22141	22157	0.00
kroC100	100	20749	20749	21669	2,2170	20749	20880	0.02	20749	20749	0.00
kroD100	100	21294	21294	22878	3,7195	21294	21374	0.04	21309	21389	0.04
eil101	101	629	629	636	0,56	629	637	0.54	630	633	0.22
bier127	127	118282	118282	122128	1.62	118282	118693	0.08	118282	118730	0.06
ch130	130	6110	6110	6394	2.32	6110	6155	0.23	6141	6174	0.42
ch150	150	6528	6528	6782	2.07	6528	6584	0.34	6528	6611	0.33
gil262	262	2378	2378	2670	6.14	2380	2410	0.08	2382	2418	0.68
A280	280	2579	2579	2783	3.96	2579	2611	0.30	2579	2592	0.51

To assess the collected results, the content of the results in the table 1 is translated into the following graph.

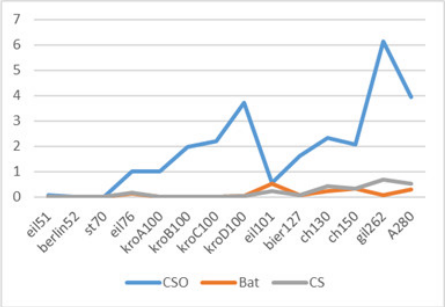


Figure 1 : RPD CSO, BA, CS to solve TSP

By analyzing the variation of the error percentage, after the application to some benchmark instance of TSPLIB, it seems clear that the BA and CS and algorithm is the best one to solve the TSP then the CSO, because the error percentage, as appears in the graphs, is lower than the CSO. Which means that the BA and CS algorithms are more efficient to solve the real life application based the TSP.

5 Conclusion

This research paper presented the application of some metaheuristics to solve the travelling salesman problem, to choose the best metaheuristic to solve some real applications based TSP problems. The studied methods are the cat swarm optimization algorithm, Bat Algorithm, Cuckoo search algorithm. This research paper aims to compare the relative percentage error by the application of these recent metaheuristics to some benchmark instances. The computational results show that the bat algorithm, and cuckoo search are more efficient than other cat swarm optimization, and had almost a similar error percentage.

In future researches, we will try to provide an application of a more performant method to solve a real application area based TSP such as the scheduling problem, Frequency Assignment Problem, applications in medical image processing.

REFERENCES

[1] N. Biggs, E. K. Lloyd and R. J. Wilson, Graph Theory, 1736-1936, Oxford University Press, 1976.

[2] M. Garey and D. S. Johnson, Computers and intractability : a Guide to the Theory of NP-Completeness Freeman San Francisco, San Francisco: Freeman, 1979.

[3] J. R. Allwright and D. Carpenter, "A distributed implementation of simulated annealing for the travelling salesman problem," Parallel Computing, vol. 10, no. 3, pp. 335-338, 1989.

[4] M. Gendreau, G. Laporte and F. Semet, "A tabu search heuristic for the undirected selective travelling salesman problem," European Journal of Operational Research, vol. 106, no. 2, pp. 539-545, 1998.

[5] M. Bouzidi and M. E. Riffi, "ADAPTATION OF THE HARMONY SEARCH ALGORITHM TO SOLVE THE TRAVELLING SALESMAN PROBLEM," Journal of Theoretical & Applied Information Technology, vol. 62, no. 1, 2014.

[6] A. Ouaarab, B. Ahiod and X.-S. Yang, "Discrete cuckoo search algorithm for the travelling salesman problem," Neural Computing and Applications, vol. 24, no. 7-8, pp. 1659-1669, 2014.

[7] Y. Nagata and D. Soler, "A new genetic algorithm for the asymmetric traveling salesman problem," Expert Systems with Applications, vol. 39, no. 10, pp. 8947-8953, 2012.

[8] K. Jun-man and Z. Yi, "Application of an improved ant colony optimization on generalized traveling salesman problem," Energy Procedia, vol. 17, pp. 319-325, 2012.

[9] X. H. Shi, Y. C. Liang, H. P. Lee, C. Lu and Q. Wang, "Particle swarm optimization-based algorithms for TSP and generalized TSP," Information Processing Letters, vol. 103, no. 5, pp. 169-176, 2007.

[10] L.-P. Wong, M. Y. H. Low and C. S. Chong, "A bee colony optimization algorithm for traveling salesman problem," in Second Asia International Conference on Modelling & Simulation, IEEE, 2008, pp. 818-823.

[11] A. Bouzidi and M. E. Riffi, "Discrete Cat Swarm Optimization to Resolve the Traveling Salesman Problem," Interneational Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 9, pp. 13--18, 2013.

[12] Y. Saji and M. E. Riffi, "A novel discrete bat algorithm for solving the travelling salesman problem," Neural Computing and Applications, pp. 1-14, 2015.

[13] K. Katayama, H. Sakamoto and H. Narihisa, "The efficiency of hybrid mutation genetic algorithm for the travelling salesman problem," Mathematical and Computer Modelling, vol. 31, no. 10, pp. 197-203, 2000.

[14] H. Duan and X. Yu, "Hybrid ant colony optimization using memetic algorithm for traveling salesman problem," in Approximate Dynamic Programming and Reinforcement Learning, 2007. ADPRL 2007. IEEE International Symposium on, IEEE, 2007, pp. 92-95.

[15] Y. Marinakis and M. Marinaki, "A hybrid multi-swarm particle swarm optimization algorithm for the probabilistic traveling salesman problem," Computers & Operations Research, vol. 37, no. 3, pp. 432--442, 2010.

[16] L.-P. Wong, M. Y. H. Low and C. S. Chong, "Bee colony optimization with local search for traveling salesman problem," International Journal on Artificial Intelligence Tools, vol. 19, no. 3, pp. 305-334, 2010.

[17] S.-C. Chu, P.-W. Tsai and J.-S. Pan, "Cat swarm optimization," in PRICAI 2006: Trends in artificial intelligence, Springer, 2006, pp. 854--858.

[18] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in Nature inspired cooperative strategies for optimization (NICSO 2010), Springer, 2010, pp. 65-74.

[19] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on, IEEE, 2009, pp. 210-214.

[20] G. Reinelt, "TSPLIB—A traveling salesman problem library," ORSA journal on computing, vol. 3, no. 4, pp. 376-384, 1991.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Creatinine, Urea and Uric Acid in Hospitalized Patients with and Without Hyperglycemia Analysis using Generalized Additive Model

Souad Bechrouri, Abdelilah Monir, Hamid Mraoui, Mohamed Choukri, Ennouamane Saalaoui
Department of Chemistry, Mohamed First University, Oujda, Morocco
choukrimohamed@hotmail.com;souad.bechrouri@gmail.com, abdelilah.monir@gmail.com,
hamid_mraoui@yahoo.fr; saalaoui_ennouamane@yahoo.fr

ABSTRACT

Hyperglycemia is an important risk factor for heart disease and premature mortality. In hospitalized patients, it is related to an increase in morbidity and development of other disease like kidney disease. To evaluate the existent relation between hyperglycemia and different biochemical parameters, we have proceeded to analyzi the difference between groups of patients which are separated according to the Glucose critical value (1.26 g/L). Generalized additive models (GAM) was used in the aim to model the relation between estimated glomerular filtration renal (eGFR) and some biochemical parameters. Our study was conducted on a data set recorded on 5600 hospitalized patients in CHU Oujda. Our statistical study revealed that the hyperglycemic patients present an increase in values of each of uric acid, creatinine, urea and triglycerides. This increase is accompanied by a loss in HDL cholesterol and eGFR. Regarding the gender of patients, results show a difference between males and females according to each of parameters: creatinine, urea, uric acid, total cholesterol, HDL and LDL cholesterol. Moreover, results show lower values of eGFR for males. The model which explain the eGFR shows a non-linear relation between dependent variable eGFR and some predictors (e.g. urea, calcium and uric acid parameters).

Keywords-component; hyperglycemia; eGFR; generalized additive model;

1 Introduction

Hyperglycemia is considered as an important risk factor for heart disease, even for levels under those currently used to diagnose diabetes. More specifically in hospitalized patients, hyperglycemia is related to an increase in morbidity, hospital stays and the risk of death [1]. Currently around 415 million in the world have diabetes and 318 million of adults have an impaired glucose tolerance (i.e. adults who have blood glucose levels higher than normal but not sufficiently to be classified as diabetics) which expose them to a high risk for developing diabetes in the future [2]. The time lived from the beginning of diabetes 2 and its discover is on average 10 years, a period in which some complications can be developed such as renal disease[3], [4]. The World Health Organization (WHO) estimates that high level blood glucose is one of the important risk factors for premature mortality. This pathology continues to increase in number and importance in parallel with Economic development and change in lifestyles characterized by a reduced physical activity and increased obesity [2]. According to the IDF atlas [2], the highest prevalence was in

DOI: 10.14738/tmlai.54.3212
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3212

2011 in the Middle East and North Africa (12.5%). Around ± 35.4 million people are living with diabetes in the Middle East and North Africa in 2015. Over 40.6 % of them are undiagnosed. The prevalence in this region is estimated at ±9.1 %. In Morocco, the prevalence of diabetes has increased to 7 % due to a change in the lifestyle and diminution of physical activity [5]. The majority of diabetics have type 2 diabetes mellitus [6].

Diabetic Kidney disease is one of the important consequence of hyperglycemia. One third of diabetic patients have this complication[7].

The main objective of this work is to study some biochemical parameters in hospitalized patients. We also aim at determining levels of uric acid, urea, creatinine, cholesterol, HDL cholesterol, LDL cholesterol and triglyceride in this population. In the same way, we investigate the correlation between these variables. In addition, we search the effect of dysglycemia and others biochemical parameters on eGFR.

2 Materials and methods

2.1 Sampling procedure

From October 2015 to Jun 2016, data were recorded from hospitalized patients at the Mohamed VI Hospital Center (CHU) in Morocco. Our data set is constructed from measures recorded on 5600 patients who visited the Hospital in the period of our study. In this work, we select and include patients who have a valid value for fasting blood glucose (FBG). We excluded all patients who have a large percentage of missing data. In the final, the population selected for this study is composed of 4011 patients (where 2025 are males). Hyperglycemia is identified by a Fasting Blood Glucose FBG value greater than 1.26 g/L. This study is conducted under anonymous and refereeing only on the biochemical characteristic, gender and age.

2.2 Data Collection

Data were recorded in the first admission of patient to one of the CHU services. The data includes demographic information as sex and age. Additionally, we have 15 laboratory tests including fasting blood glucose, creatinine, urea, uric acid, fasting blood total cholesterol, triglyceride and others. Blood samples were analyzed in the CHU for all laboratory tests. The definition of hypercreatininemia, hyperuremia, hyperuricemia, hypercholesterolemia, hyperHDLemia, hyperLDLemia and hypertriglycerimia is based on the rules below: plasma glucose: for both males and females 0.70 - 1.10 g/L, plasma uric acid: Males 30 à 70 mg/L, Females 25 à 60 mg/L; plasma creatinine: Males 6 à 12 mg/L, Females 5 à 10 mg/L; plasma urea: for both males and females 0,15 à 0,45 g/L; plasma triglyceride: 0,45 à 1,50 g/L ; plasma total cholesterol : 1,50 à 2 g/L; plasma HDL cholesterol: 0,50 à 0,70 g/L and plasma LDL cholesterol: 1 à 1,6 g/L.

2.3 Statistical analysis

Data were analyzed under the statistical R software, version 3.2.3. The Wilcoxon test (respectively the Student t-test) were used for mean's comparison for non-Gaussian (and Gaussian) variables. For nominal variables, Fisher's exact test was used for determining differences between groups. The correlation coefficient were analyzed by Spearman correlation analysis. The Glomerular filtration rate (eGFR) is estimated by Modification of Diet in Renal Disease MDRD 2006 function as presented in Eq. (1):

$$175 \times SCr^{-1.154} \times Age^{-0.203} \times 0.742 \; if \; female \quad (1)$$

where SCr is the Serum Creatinine level.

To search the impact of Fasting Blood Glucose and other biochemical parameters on estimated GFR, we request the use of generalized additive models [8]. The model to estimate is:

$$E(Y|X_1, \ldots, X_p) = f_0 + \sum_{i=1}^{p} f_i(X_i) \quad (2)$$

where Y represents the estimated GFR, X_i are the different predictors and f_i is an unspecified function. The last one is estimated by natural cubic spline.

Natural cubic splines are piecewise cubic polynomial functions with first and second continuous derivatives at the knots (break points) and zero first and second derivative functions at the boundary knots.

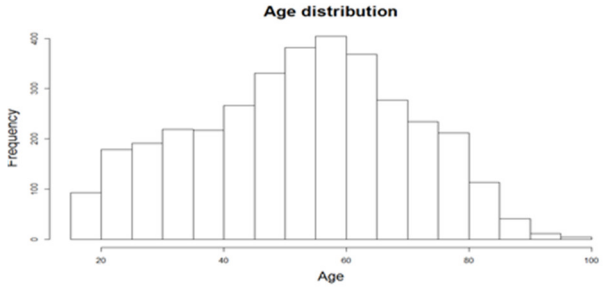


Figure 1. Participants distribution age

Table I. Biochemical Characteristics in the Population of Study

Characteristics	Total population	glycaemia >=1.26	glycaemia <1.26	p_value
N (%)	3542	1107(31.25%)	2435(68.75 %)	
Age (years)	52.99 (±17.38)	58.81(±15.61)	50.35(±17.51)	<0.0001
Male (%)	1763 (49.77%)	569 (51.92 %)	1187(48.75%)	0.1535
Glucose (g/L)	1.29(±0.68)	1.98(±0.82)	0.93(±0.17)	<0.0001
Creatinine (mg/L)	14.87(±24.21)	18.9(±27.96)	12.95(±21.75)	<0.0001
Uric acid (mg/L)	56.11(±25.53)	61.39(±29.85)	53.54(±22.73)	<0.0001
urea (g/L)	0.53(±1.08)	0.65(±0.56)	0.47(±1.23)	<0.0001
total cholesterol (g/L)	1.72(±0.58)	1.75(±0.64)	1.71(±0.56)	0.445
HDL cholesterol (g/L)	0.41(±0.15)	0.38(±0.14)	0.42(±0.15)	<0.0001
LDL cholesterol (g/L)	1.04(±0.44)	1.04(±0.48)	1.04(±0.42)	0.4983
triglyceride (g/L)	1.39(±1.05)	1.65(±1.26)	1.27(±0.91)	<0.0001
eGFR (ml/min . 1.73 m²)	101.4(46.32)	84.43(45.11)	100.64(46.58)	<0.0001

where SCr is the Serum Creatinine level

3 Results

3542 patients aged between 18 and 96 years were included in this study and were hospitalized in CHU (Fig 1). In average, the age of this population is 52.99 y (±17.38 y). 50.22% of them are females (1779) and 49.77 % are males (1763). Their mean age is 53.05 y (±17.12 y) and 52.93 y (±17.65 y) respectively. 1096 patients have a fasting blood glucose upper than 1.26 g/L. The mean age for this sub-population is 58.84 y (±15.52 y). 569 (51.92 %) of them are males and 527 (48.08 %) females. They are aged in mean 57.88 y (±15.92 y) and 59.88 y (±15.01 y) respectively. 408 patients who have plasma glucose between 1.10 g/L and 1.26 g/L are aged 55.09 y (±16.76 y).

Table I summarizes the baseline biochemical analysis by hyperglycemia status. Patients with hyperglycemia were significantly older than those who have a plasma glucose level under than 1.26 g/L (58.81y (±15.61 y) vs 50.35 y (±17.51 y) respectively). The two groups have a significant difference between plasma levels of glucose, uric acid, urea and creatinine. These values are more important in hyperglycemic group. HypoHDLemia was more important in the hyperglycemic group. The total cholesterol and LDL cholesterol are non-significantly different between the two groups. However, for triglyceride levels, it was greater in the first group with 1.62 g/L ((±1.26 g/L). All these differences were significant at 95 % confidence level. In addition, the GFR value estimated in hyperglycemic group is significantly less than this estimated in the other group ($p<0.0001$).

The correlation coefficients between different measured parameters are presented in table II. The Age is significantly correlated with creatinine, uric acid and urea ($p<0.005$). Also, the FBS is positively related with creatinine, urea and negatively related with eGFR. The correlation is significantly negative between the creatinine and different forms of cholesterol and significantly positive between creatinine and both urea, uric acid. Urea and uric acid are strongly correlated with each other and with GFR.

Table 2 Correlation Table Between Measured Variables in Hyperglycemic Group

	Age	FBS	Creatinine	Uric acid	Urea	Total cholesterol	HDL cholesterol	LDL cholesterol
Age	1							
FBS	0.02	1						
Creatinine	0.24***	0.17***	1					
Uric acid	0.15**	0.05	0.69***	1				
Urea	0.25***	0.14***	0.78***	0.68***	1			
Total cholesterol	-0.08	0.05	-0.16***	-0.10	-0.17***	1		
HDL cholesterol	0.03	-0.04	-0.18***	-0.18**	-0.19***	0.48***	1	
LDL cholesterol	-0.06	0.02	-0.17***	-0.11	-0.16**	0.91***	0.38***	1
Triglyceride	-0.02	0.11*	0.06	0.15**	0.03	0.46***	-0.17***	0.28***
eGFR	-0.36***	-0.18***	-0.92***	-0.69***	-0.75***	0.12**	0.10*	0.11*

Significance levels: *: <0.05, **: <0.005, ***: <0.0005

Table III represents a comparison of various measured variables according to the gender. Females presented a high fasting glucose level (2.03g/L (±0.92)) than males (1.92 g/L (±0.73)). Also, it seems that they have a higher level of plasma cholesterol (1.84 g/L (±0.67) vs 1.67 g/L (±0.60)) and LDL cholesterol (0.98 g/L (0.46) in males and 1 g/L (0.49) in females). Furthermore, males seem have a hypoHDLemia compared with females. These differences were significant at 95 % confidence level.

Table 3. Biochemical parameters by gender in hyperglycemic patients

Characteristics	Males (N=560)	Females(N=531)	p_value
Age (years)	57.92(15.98)	59.77(15.15)	0.9824
Glucose (g/L)	1.92(0.73)	2.03(0.89)	0.01012
Creatinine (mg/L)	21.29(32.15)	16.25(22.31)	<0.0001
Uric acid (mg/L)	64.29(32.55)	58.51(26.69)	<0.0001
Urea (g/L)	0.71(0.72)	0.58(0.57)	<0.0001
Total cholesterol (g/L)	1.67(0.60)	1.84(0.67)	<0.0001
HDL cholesterol (g/L)	0.35(0.13)	0.42(0.14)	<0.0001
LDL cholesterol (g/L)	0.98(0.46)	1(0.49)	<0.0001
Triglyceride (g/L)	1.62(1.01)	1.68(1.47)	0.6213
eGFR (ml/min . 1.73 m²)	77.48(42.15)	91.93(47.00)	<0.0001

To investigate the relationship between eGFR and biochemical parameters, we use the generalized additive model below:

$$eGFR \sim \alpha_0 + \alpha_1 f_1(glucose) + \alpha_2 f_2(urea) + \alpha_3 f_3(calcium) + \alpha_4 f_4(potassium) + \alpha_5 f_5(triglyceride) \quad (3)$$

These variables are selected by a stepwise method. The first model is a model without transformation, a generalized linear model, whereas for the second model,f_i are estimated by spline functions.

The first one explains 64% of the deviance. If we take this model with uric acid instead of urea, we have 44.5 % of explained deviance. These two variables can't be used in the same model because of its collinearity. In table V we present results of generalized linear model. Males have a greater risk for developing a kidney disease compared to females. Urea has an important relation with the eGFR. Moreover, an increase of one unit in glucose level implies a decrease in GFR value multiplied by 8.71. The same remark can be done for other parameters.

The model with uric acid is summarized in table VI. The triglyceride is at the limit of significance at risk level of 5%. Also, the calcium has a positive effect on eGFR in this model.

The use of cubic spline allowed to identify a non-linear relation between calcium, urea, triglyceride and eGFR. The second model explains 77.2 % of the deviance while the first one explains only 64% of the deviance. The adjusted coefficient R is equal to 0.64 in the simple model and 0.76 for natural cubic spline model. Figure 2 shows the relationship between different predictors and GFR.

According to table VII, there is no evidence of nonlinearity for glucose variable, so we decide to include it as linear term. For other parameters, we have choose natural spline because it impose the constraint of linearity in outside of the tails. We remark that other parameters like potassium, calcium and triglyceride don't have a nonlinear effect on eGFR. The model obtained in this case is:

Table 5. Variables effect on estimated GFR with urea

PARAMETERS	Estimate	Se	T value	P(>| t |)
Intercept	205.5209	17.2704	11.900	< 2e-16
Urea	-42.9316	2.1268	-20.186	< 2e-16
potassium	-10.6904	2.5095	-4.260	2.53e-05
Calcium	-0.4480	0.1672	-2.680	0.007666
glucose	-8.7137	2.6178	-3.329	0.000951
triglyceride	-4.1004	1.5613	-2.626	0.008952

$$DFG \sim \alpha_0 + \alpha_1 \sum_{i=1}^{4} B_i(urea) + \alpha_2 * glucose + \alpha_3 \sum_{i=1}^{3} B_i(potassium) + \alpha_4 \sum_{i=1}^{3} B_i(calcium) + \alpha_5 \sum_{i=1}^{3} B_i(triglyceride)$$

Iii. Dicusssion and Conclusion

Table 6. Variables effect on estimated GFR with uric acid

PARAMETERS	Estimate	Se	T value	P(>| t |)
Intercept	184.07019	21.64776	8.503	3.38e-16
Uric acid	-0.71797	0.06854	-10.475	< 2e-16
potassium	-23.60241	2.97037	-7.946	1.83e-14
Calcium	0.59921	0.19252	3.113	0.00198
glucose	-13.97445	3.25327	-4.296	2.17e-05
triglyceride	-3.83248	1.95915	-1.956	0.05111

Table 7. Non linear effect of biochemical parameters on GFR

Parameters	Estimate	Se	T value	P(>| t |)
Intercept	149.95624	14.73084	10.180	< 2e-16
ns(urea, 4)1	-73.19837	6.54832	-11.178	< 2e-16
ns(urea,4)2	-126.55307	8.42283	-15.025	< 2e-16
ns(urea,4)3	-130.65652	13.55126	-9.642	< 2e-16
ns(urea,4)4	-107.51349	11.29973	-9.515	< 2e-16
glucose	-6.51121	8.17356	-0.797	0.42614
ns(potassium, 3)1	-17.20104	5.89575	-2.918	0.00372
ns(potassium, 3)2	0.08205	18.44900	0.004	0.99645
ns(potassium, 3)3	2.73235	12.55371	0.218	0.82781
ns(calcium, 3)1	-19.01104	6.82003	-2.788	0.00556
ns(calcium, 3)2	-36.55901	22.73195	-1.608	0.10855
ns(calcium, 3)3	-29.41872	11.55790	-2.545	0.01129
ns(triglyceride, 3)1	-1.57009	8.22592	-0.191	0.84872
ns(triglyceride, 3)2	-30.63209	14.91809	-2.053	0.04068
ns(triglyceride, 3)3	-39.90629	18.51749	-2.155	0.03174

Our studied population is homogeneous in its composition. We have a quasi-equal proportion of males and females. The hyperglycemic group was older than normal group and they have no difference in the proportion of Males. Examination of the biochemical profile of our population reveals the existence of

hyperglycemia in the one third of the studied patients. It is less than the percentage of this founded in [9]. Also, it divulge the existence of a metabolic disorder which is characterized by a dyslipidemia, hypercretininemia, hyperuricemia and hyperuremia in hyperglycemic group [10]. Precisely, patients with a higher blood glucose have a decrease in HDL cholesterol and increased triglyceride levels [11]. A higher value of uric acid induce a low eGFR so it can inform about renal function [12].

In the additive generalized model, we have eliminated variables which have a strong correlation such as uric acid. If we choose this variable instead of urea, we obtained only 44.5 % of the explained deviance. This study has revealed a non-linear relation between eGFR and urea on the one hand, and between eGFR – calcium in the other hand. Linear relation between hyperuricemia and eGFR was demonstrated in previous works[13]. Urea is used as a initial diagnosis of chronic kidney disease.

Some limits of this study, it was the fact to have a lot of missing values in our data set. So for future work, we will impute missing data. Also, this study can be extend in the future to other CHU in Morocco. It could be interesting to apply this study on the longitudinal data and to perform it with a local regression.

REFERENCES

[1] E. H. Baker, C. H. Janaway, B. J. Philips, A. L. Brennan, D. L. Baines, D. M. Wood, and P. W. Jones, “Hyperglycaemia is associated with poor outcomes in patients admitted to hospital with acute exacerbations of chronic obstructive pulmonary disease.,” Thorax, vol. 61, no. 4, pp. 284–9, Apr. 2006.

[2] IDF, Diabetes. 2015.

[3] M. I. Harris, R. Klein, T. A. Welborn, and M. W. Knuiman, “Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis.” Diabetes Care, vol. 15, no. 7, pp. 815–9, Jul. 1992.

[4] C. S. Levetan, M. Passaro, K. Jablonski, M. Kass, and R. E. Ratner, “Unrecognized diabetes among hospitalized patients,” Diabetes Care, vol. 21, no. 2, pp. 246–9, Feb. 1998.

[5] M. A. Tazi, S. Abir-Khalil, N. Chaouki, S. Cherqaoui, F. Lahmouz, J. E. Sraïri, and J. Mahjour, “Prevalence of the main cardiovascular risk factors in Morocco: results of a National Survey, 2000.,” J. Hypertens., vol. 21, no. 5, pp. 897–903, May 2003.

[6] V. Kumar, A. K. Abbas, N. Fausto, S. L. (Stanley L. Robbins, and R. S. Cotran, Robbins and Cotran pathologic basis of disease. Elsevier Saunders, 2005.

[7] . A. Roberto Pecoits-Filho1*, Carolina C. R. Betônico3, Alisson Diego Machado2, Erika B. Parente4, Márcia Queiroz2, João Eduardo Nunes Salles4, Silvia Titan2, and and Sergio Vencio, “Interactions between kidney disease and diabetes: dangerous liaisons,” Diabetol. Metab. Syndr., 2016.

[8] T. Hastie and R. J. Tibshirani, Beneralized Additive Models. Chapman & Hall, 1990.

[9] S. Feldman-Billard, M. Joubert, R. Morello, F. Dorey, D. Seret-Begue, F. Getin-Bouyer, P. Jan, A. Colobert, E. Verlet, M. Roques, and Y. Reznik, “High prevalence of diabetes mellitus and hospital-related hyperglycaemia in French general wards,” Diabetes Metab., vol. 39, no. 5, pp. 454–458, 2013.

[10] E. B. Sellam and A. Bour, "Prévalence des facteurs de risque de maladies cardiovasculaires chez la femme à Oujda (Maroc)," Médecine des Mal. Métaboliques, vol. 10, no. 1, pp. 63–69, 2016.

[11] A. D. Mooradian, "Dyslipidemia in type 2 diabetes mellitus," Nat. Clin. Pract. Endocrinol. Metab, vol. 5, no. 3, pp. 150–159, Mar. 2009.

[12] B. SATIRAPOJ, O. SUPASYNDH, A. CHAIPRASERT, P. RUANGKANCHANASETR, I. KANJANAKUL, D. PHULSUKSOMBUTI, D. UTAINAM, and P. CHOOVICHIAN, "Relationship between serum uric acid levels with chronic kidney disease in a Southeast Asian population," Nephrology, vol. 15, no. 2, pp. 253–258, Mar. 2010.

[13] Y. Xu, X. Liu, X. Sun, and Y. Wang, "The impact of serum uric acid on the natural history of glomerular filtration rate: a retrospective study in the general population." PeerJ, vol. 4, p. e1859, 2016.

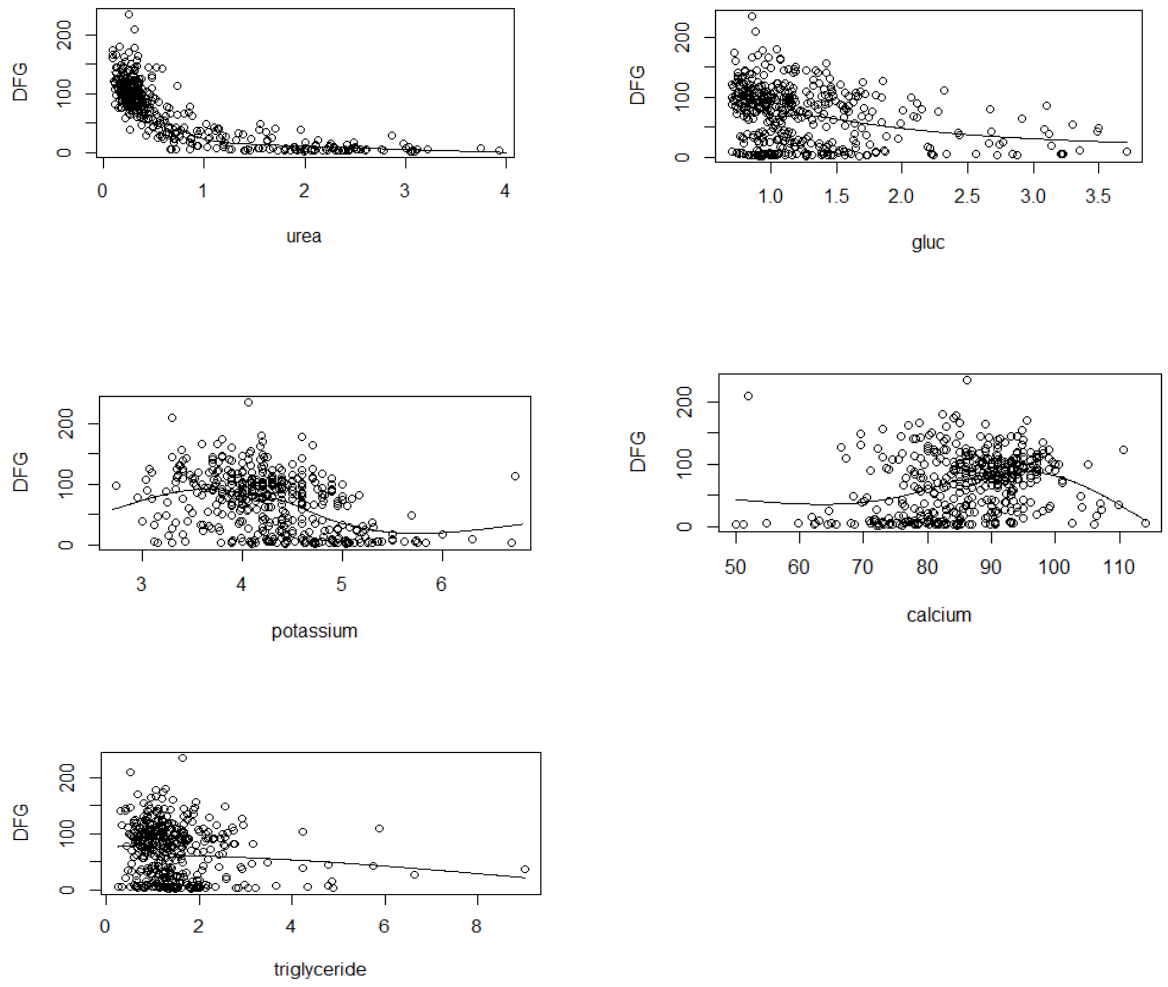


Figure 2. eGFR versus urea, glucose, potassium, calcium and triglyceride with natural spline adjustment by generalized additive regression

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

New Cloud Economy For Resolving A Problem Cloud Computing Adoption

M. Elmahouti, O. Achandair, S. Khoulji, M. L. Kerkeb,
Information Systems Engineering Research Group, Dept. Telecommunications.
National School of Applied Science of Tetouan,
Abdelmalek Essaadi University, Tetuan, Morocco
mouradelmahouti@gmail.com, o.achandair@gmail.com, khouljisamira@gmail.com,
kerkebml@gmail.com,

ABSTRACT

Cloud computing , IT still called to order, it is a model that allows access to information resources on demand, we can consider it as a unique Technology to satisfy needs and respond to customer demands by guaranteeing an efficient quality of serviceIT management.

This paper gives an analysis of cloud computing along its components, also the internal and external institutional factors that influence the adoption of cloud computing in the majority of enterprises.

It aims to be innovative by answering these questions of allocation and availability of network services in an intra-datacenter environment by implementing solutions based on new technologies that is OpenStack[1].

The research problem addressed in this paper is the proposal of a new architecture and optimized algorithms of continuity and availability of the service. These will help identify the deployment of existing vulnerabilities. On the other hand, offer a very high level of availability, security and quality of service.

Keywords—Cloud computing, Private cloud, Public cloud Information resources, Information security, OpenStack

1 Introduction

Today cloud computing is one of the information technologies that shows an intrinsic transformation for improving the quality of service rendered. The goal is to minimize the reflection time on the processing of an exponential amount of dynamic data. Today, it is difficult to cope with the increasing difficulties for a good functioning of the different architectures, the new concept cloud computing provides a fascinating proposition to the companies to outsource the IT infrastructures.

Generally, four resources can be used through the Internet and by providing an extended economy, starting with the infrastructure system like storage, the second type is the software system that supports the middleware part or other parts, and the application system as is the case of GOOGLE, finally the commercial procedures.

DOI: 10.14738/tmlai.54.3213
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3213

This article describes the design of cloud computing, the design of the architecture adopted, the points gained from this model and an analysis of the major data technologies, as well as an understanding of the factors that hamper the adoption of this model.

The article also responds to the concerns expressed today by most information systems managers who have not been able to advance in the adoption of this new cloud computing technology because they have recently invested very large budgets for their current platforms, presenting them with an intelligent solution that meets future needs with a saving of budgets reserved for the IT part, the solution is based on the free software OpenStack[3]., which consists in setting up a flexible cloud "Private- Public". A private cloud that supports their current platforms and a public cloud that manages their loads when exceeding the company's internal technical capabilities by ensuring a high level of data exchange security between the two clouds.

The article gives a presentation on the architecture, synopsis and types of cloud computing, discussing the benefits of adopting this technology, and the internal and external factors that hampered the rapid adoption of the cloud by most Presenting an effective technical solution to meet the constraints encountered.

2 Cloud computing : Synopsis

Cloud computing is a computer system based on the use of the Internet to exploit IT resources[2]., these resources are used through a network approach, connecting to a pool that shares configurable resources such as servers, storage, networks And applications that can be directly assigned to customers by service providers.

Cloud computing consists of some basic aspects:

- Measured service: The IT resources are shared by various customers, ensuring a measure of usage of each, with control and monitoring, which gives clarity to both parties: consumer and supplier.
- Resource pooling: The provider of cloud computing makes available to its various clients resources in common using the multi - support mode to share the various virtual and physical resources with a possibility of renewal according to the demand of the client[4].
- Rapid elasticity: Great flexibility in the release and allocation of resources, real-time adjustment, with a perfect control of the consumption of resources.
- On-demandself-service: The existence of an application that responds spontaneously to allocate resources automatically without going through a human intervention.

2.1 Architectureof Cloud Computing

The architecture of cloud computing[5]. can be divided into four layers, starting with the application layer, the layer that forms the platform, the infrastructure layer of the party and the hardware layer as shown in “Figure 1”.

- Application layer : This is the top level of cloud model composed specific cloud applications, while different from conventional applications, it will secure the automatic scale part to get great performance with reduced costs.
- Platform layer: It is the secondary layer that forms the operating structure, whose role is to reduce

the load by extending the application directly to the storage VM.

- Virtualization layer: Called the infrastructure layer that remains fundamental to the cloud system, which ensures a dynamic aspect of the various components based on KVM virtualization technologies such as the storage pool
- Hardware layer: This is the bottom layer of the architecture of the cloud, responsible for maintaining the physical cloud resources, such as routers, cooling systems and power , this layer contains a number of interconnected servers switches and other devices.

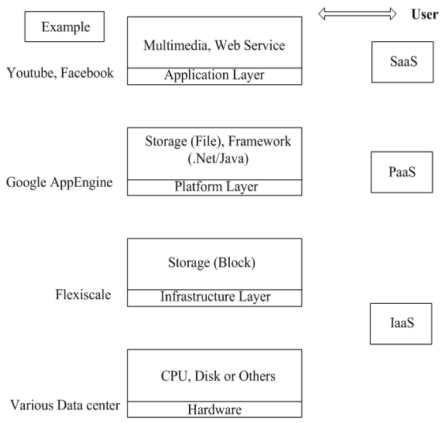


Figure1: Architectural View of Cloud Computing

2.2 Cloud Hierarchy

In most scenarios, the Cloud hierarchy is provided by four levels of users, first level users, field staff who are the first users who coordinate basic services, then transaction authors, and finally developers (CI), See “Figure 2”.

- User:These are the end users, in this layer we even find indirect users like decision-makers on policies to choose, analysis and illustration.
- Service integration and composition: At this level, the user judged expert must be able to focus on training results intended for an end user can integrate and manage existing services, able to customize and provide updated services and inspect the use of services and collect service information.
- Service author: the authors are the developers of distinct bases of services that can be used directly or combined in other aggregates, and complex service systems by experts in providing the service. The authors should be experts components having a positive outlook needs and categories to ensure such registration services, license management tools, fault tolerance.
- Developers: Cyber-infrastructure (CI) developers are responsible for developing cloud structure , be experts is an asset in the areas of storage , middleware, networks, generation systems , creating devices for services , scheduling algorithms and others[6].

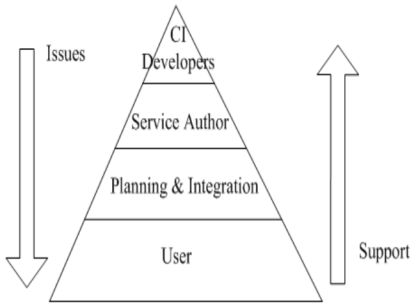


Figure 2: The Hierarchy of Cloud Computing

2.3 Service Model of Cloud Computing

Cloud management is completely centralized. The database server has the role to specify the balance between customer demand and system capabilities, and monitor network traffic.

Users achieve IT infrastructures from clouds system and next run applications inside thus Clouds computing distribute customer along services to approach software, infrastructure and data resources as shown in “Figure 3”.

- Infrastructure as a Service: IaaS Infrastructure as a Service is one of three main categories of cloud computing services, this service gives access to computing resources in a virtualized environment, the "Cloud", through a public connection, typically the Internet. For IaaS, the resource is supplied hardware virtualized or, in other words , an IT infrastructure. The service definition includes offerings such as server space, network connections, bandwidth , IP addresses and load balancers . Physically, the hardware resources come from a multitude of servers and networks typically distributed across many data centers, the cloud provider’s responsibility to maintain. Meanwhile, access to virtualized components is given to the client company so that it can build its own IT platforms. [7].
- Platform as a Service: often simply referred to as PaaS, is a category of cloud computing that provides a platform and environment to allow developers to build applications and services over the internet. PaaS services are hosted in the cloud and accessed by users simply via their web browser.
- Software as a Service: Any cloud service that allows customers to have access to software applications via the Internet or Internet as if applications hosted in a service provider. These applications are hosted in the cloud and can be used if they have permissions, both by individuals and by organizations. Google, Twitter, and Facebook are all examples of SaaS, or users can connect to the services by any device with an Internet connection.

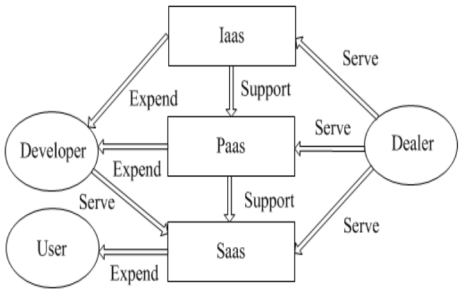


Figure 3: Abstractions of Cloud Computing

3 Types of cloud

There are different types of cloud that allow a wide choice for users in the desired formula, namely the private cloud, public cloud, hybrid cloud and virtual private cloud.

- Private cloud: this model is mainly chosen to provide the client more control and privacy on his cloud based environment, which consist of offering a pool of physical computing resources as a service accessible through a virtualized environment.
- Public cloud: This type of cloud is to make available a shared platform to serve an unlimited number of users via the Internet. Also, this means optimal use of virtualized physical resources, and therefore a consumption model based on 'you pay as yougo'. The entirely infrastructure that host client environment belong to the cloud provider, who is responsible of all the management and maintenance layers[8].
- Hybrid cloud: it is about coexisting and communicating a private cloud and a public cloud. This type of cloud is often used for the purpose of rise-time load as permitted by public cloud. The only difference is that in this case it would be linked to a private or internal cloud for communicating the two infrastructures.

Virtual private cloud: This model is a hybrid cloud environment provided by the public cloud; the cloud provider guarantees an access to an isolated portion of his infrastructure for private use. Virtual private cloud is offered through a secure connection between the enterprise and the provider. The entire infrastructure belong the provider who is responsible for all the public cloud operations in addition to isolating the concerned portion[9].

Table I: Comparison of various representative cloud platforms

Feature	Organization		
	Google	Amazon	Microsoft
Focus	Platform	Infrastructure	Infrastructure
Business Type	Web Application	Storage, Compute	Storage
Virtualisation	Application container	Xen hypervisor	OS level
Web APIs	Yes	Yes	Obscure
Compute Scheme	Python	Linux AMI	Not applicable

4 The benefits to adopt cloud computing for other data analysis technologies. (Example: BigData)

Big data background:

The exponential growth of data, generated continuously in different formats and from different sources, has lead traditional architecture and infrastructures to face many limitations to store and process such data sets, organizations are getting more and more convinced that the cloud computing is the right structure to handle their big data projects. Big data technologies start when traditional database management and traditional processing applications are facing limits and difficulties to perform[10].

In the following, we are going to describe important big data characteristics and cloud major benefit for big data.

Gartner and many IT organizations used the so called '3V' model to describe big data, which stands for :

- Volume: huge amount of data are generated from different sources such as social media, Internet of Things (IoT), mails, purchase transaction records and many others. It is estimated that 90% of the existing data has been generated during the last two years according to a new infographic compiled by Singapore- and India-based Aureus Analytics. Storing such data sets is actually challenging the researches, it need a highly scalable and reliable storage withminimal costs. Traditional systems are still needed like SAN (storage area network), still too expensive, but not designed to handle this volume/types of data, also the remaining time to query the data from it could presents many issues while performing complex computing. Most of the big IT players rely on Hadoop for storing the data in a distributed file system mechanism, which allow the storage to scale up to massive amount [11]..
- Velocity: Data is generated at increasing speeds, accessing all of it is not usually enough. People need to access the data at the right time [12].. While Hadoop map/reduce is more suited for batch-processing data, the high velocity data is processed with other real time analytics engine that complete the Hadoop ecosystem, essentially:

Table II: Comparison of various representative cloud platforms

package	Use case	Used by companies
Storm	high-speed, event processing system that allows for incremental computations	Twitter, Yahoo!,
Spark	must have stateful computations and exactly-once delivery; doesn't mind higher latency	Yahoo!, NASA JPL, eBay
Samza	has a large amount of state to work with	LinkedIn, Yahoo!, Metamarkets

- Variety: it is about collecting data in different formats from different sources, structured and unstructured, good and bad mixed together. in our networked world, systems are connected

every day in different level of dependencies. enterprises needs to track a large amount of systems that are getting upgraded or changed continuously, managing and modeling thousands of data-type to be filtered and analyzed is a big challenge for industry of big data especially when its coupled with the previous problems of volume and velocity[13].

- Cloud Convenience for big data: The Cloud computing offer many benefits for big data industry, principally as an infrastructure that support large-scale physical resources, networking, computing, high capacity of storage, high availability, high reliability… In parallel of that, its model is too attractive because it offers companies the possibility to reduce business costs, related to the investment and maintenance of their IT infrastructure. The maturity of the public cloud model driven by big IT players as Google, Amazon and Rackspace have pushed many enterprises to migrate toward cloud platforms[14].

5 External and Internal Institutional Factors that Influence the Adoption of Cloud Computing

This section describes the main factors that have influenced the speed of adoption of cloud computing technology by a large majority of companies, according to a survey of 100 heads of information systems in Morocco. Are classified into two families, external and internal factors:

5.1 External Factors:

The external factors that inhibit the adoption of this new mode of service are the following:

- Absence of regulatory organizations
- Service providers
- The media influence
- Managing change and corporate culture.
- The lack of clear legislative law in the event of breach or termination of contract between the two parties concerned: customers and service providers.
- The lack of security especially on the confidentiality of the data is the first brake enunciated by the potential DSI.
- A market that demands to be monitored to stop falling into the serious drifts that may be the resale of personal information after being pirated.
- The complexity and lack of integration for cloud adoption by companies, according to a new pan-EMEA study by Oracle. Many companies have a bad approach to cloud computing in computing and business.
- Animportant aspect of the problem is that most companies continue to finance their IT investments without aligning them with potential revenues and innovative projects: two out of three decision-makers say that IT funding is too traditional and hinders innovation: It's time to change funding models.
- Companies with an established position, such as major software companies, infrastructure providers, companies that provide services, etc. Cloud computing challenges their business model. They know they have to do it, however, much of their revenue is not realized by the cloud.
- Internet networks play a very important role in the adoption of "cloud computing" technology, telephone operators must improve the current quality of the supplied data that does not respond properly to the needs desired by the companies.

- Lack of efficient data centers.

5.2 Internal Factors

According to an in-house research of companies, exactly in the IT departments, there are several concerns that slow down the speed of adoption of the new technology of Cloud Computing, to quote exactly:

- According to a local study, the age pyramid of IT decision-makers in Morocco is between 48 and 55 years, which makes it difficult to adopt, and in particular decision-making migration to thistechnology, which requires know-how and a large capacity To ensure control while there is a great lack of knowledge of this type of services.
- Concerns about reducing the number of positions in the IT departments and the loss of current positions of responsibility following the organizational changes that will accompany the adoption of Cloud Computing.
- The immediate ICT environment within their company blocks progress.
- Current equipment has cost too much to be replaced so quickly
- The complexity of network architectures already present in companies
- There is also mistrust on the part of IT managers who are afraid of losing power or even seeing their leadership disappears.
- IT managers believe that the adoption of the cloud will slow down the career development of their employees, IT managers will only have to interface the internal demands of the business with Cloud service providers.

6 Private-Public Cloud Model based on OpenStack

To encourage companies to adopt this new cloud computing technology, minimizing budgets in the future and taking advantage of the state of the existing and existing platforms, our proposal is to create two models "Private : Cloud A"And" public : Cloud B "cloud based on a free solution that is the OpenStack[15].

Our approach is to create a "private - public" cloud: a private cloud (Cloud A) that supports the existing platform, and once the number of requests exceeds the capacity of the private platform, these requests will be refused internally by routing them to the Cloud to satisfy them by the public cloud (Cloud B) by paying just the rental costs, And once the internal resources are released, automatically some of the applications running on the public cloud will be supported again by the private cloud by minimizing rental costs[16].

6.1 Architecture proposed for the cloud "Private - Public"

The “Figure 4”shows the mechanism to be used by proposing the three algorithms to follow in order to pass the two-way passage between the two clouds, private and public, to ensure this proposition:

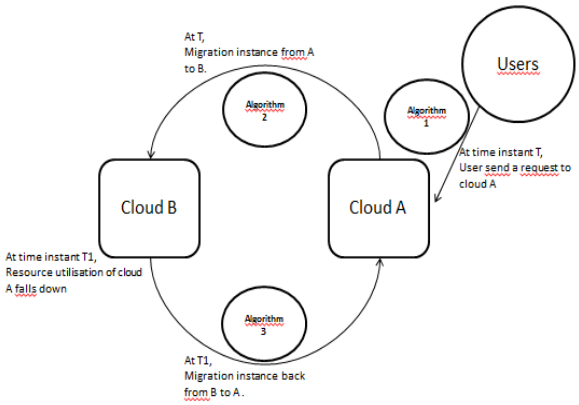


Figure4: Cloud integration

6.2 Presentation of the three algorithms used:

The main brick of Openstack is Nova. Its purpose is to manage the resources of calculation of the infrastructure used, using the command NOVA BOOT to attack our platform, with its syntax below:

nova boot --image imageID --flavor flavorID --nic net-id=nicID

imageID :To select the operating system.

net-id=nicID : To tell the cloud which subnet is used for an instance.

flavorID :Reflects memory, disk and virtual process requirements.

To check the actual load on the cloud A, an implementation of the load analysis function is done, the functioning of the "analyze load" function will be explained in the algorithm 1 below[16].

Once the "analyze load" function is called, all information from all compute nodes of a controller on cloud A is received, if the load on the A cloud is below the configurable threshold, the instance will be created On cloud A, but if the load is greater than the threshold, the instance will be created on the cloud B[17].

Algorithm 1 Analyzing the load

Input Parameter:

flavorid: flavor id of requested instance

Output Parameter:

val: 0 if cloud is overloaded, 1 otherwise.

procedureAnalyzing the load (f lavor id)

Oracle database connection is made to the nova database.

Extract user's requirements using flavor id from the instance types table.

totalmemory for all compute and total disk for all compute is initialized to 0.

freememory for all compute and free disk for all compute is initialized to 0.

for each compute node c do

Find the total memory and the free memory of c from the compute nodes table.

Find the total disk space and the free disk space of c from the compute nodes table.

totalmemory for all compute =total memory for all compute + total memory of c.

total disk for all compute =total disk for all c-ompute + total disk of c.

freememory for all compute =free memory for all c-ompute + free memory of c.

free disk for all compute =free disk for all co-mpute + free disk of c.

end for

if (total disk for all compute * disk threshold percentage>= users disk requirement + (to-tal disk for all compute-free disk for all

compute) then

if (total memory for all compute * ram threshold percentage) >= users memory requirement + (total memory for all compute-free memory for all compute) then

Return 1

end if

end if

Return 0

end procedure

Note that setting up an instance on another cloud requires some information on the Cloud B: Image-ID, Network-ID, which is used by algorithm 2 to position the instance on the cloud B[18].

Algorithm 2 Position the instance on the cloud B

Input parameter:

Network id: Subnet to start instance. Image id: OS launching an instance.

Tasting Id: Specify required resources

Instance Name: The name of the instance to be launched.

Procedure INSTANCE positioning ()

When running the nova boot command, the function creation ()

calls up the load to check the load

If Not exceeding threshold then

The instance is created on cloud A.

Other

It calls the createInstance () function. // Creating the instance on the cloud B

CreateInstance () uses the new startup parameters

We call the function creation () and extract the

Id of the image for the cloud B.

It Execute the NOVA command with the new parameters.

This nova boot command is executed

On the B cloud using the SSH connection.

All instance information is stored in a file, retrieved later, and stored again in the array of migrating instances to the NOVA database

End if

End of procedure.

After each migration to the cloud B, the load on the cloud A is verified, once the load is below the configurable threshold, a remigration is made again to the cloud A, the instances that will be migrated back to the cloud A follow A FIFO order, algorithm 3 below explains the function of the remigration operation:

Algorithm 3 Remigration

Input Parameter:

Migratedinstancestable: Table regrouping the instances of remigration

procedure REMIGRATION( )

The Analyzing the load() for remigration is Performed at

Of the parameterized intervals.

if the resource utilization falls below a configurable

low threshold then

do

It finds the instance to be remigrated from the

migratedinstances table using the FIFO : rule.

It launches a new instance on cloud A using

the information stored in the migrated instances table.

It copies the disk image of the migrated

instance from the cloud B to the newly launched instance on the cloud A, to: restore the current state.

It finished the remigrated instance from thecloud B to free resources.

It cleans the entry of the table of instances that have been migrated.

whileResource usage is below the migration threshold and the migrated instance table is not empty

end if

end procedure

The communication between the two clouds (Cloud A and Cloud B) is secured with the use of an SSH public key, as noted in algorithm 2.

7 Results

Currently, OpenStack installations have been deployed in the VEOLIA data center (Pilot Site: Gabon) on two dell poweredge 2950 servers. The first server acts as a private cloud (Cloud A) and the second as a public cloud (Cloud B). The memory of each server is: 1.8 TB, which acts as one of the compute nodes. 30 machines were assigned which function as calculation nodes for the two clouds, respectively. Knowing that the two clouds belong to a single subnet[19].

The three proposed algorithms are already tested, we manage to control the load on the private cloud in-house (Cloud A), once the configurable threshold is exceeded, the new instances launched at the company's breasts are migrated and directly supported To the public cloud (Cloud B), with a check every 30 seconds of the capabilities of the cloud A, once the load falls below our internally configurable threshold, the instances launched on the cloud B are again remigrated and retrieved by the cloud A by following a FIFO (First In First Out) recovery order, which will also minimize our rental costs at the public cloud provider[20].

The next step will be to generalize the study across our platform by creating a private cloud that contains the entire existing infrastructure, with the proposal of a powerful test server that will play the role of the public cloud, Sharing of all the screenshots of both implementations will be ensured to facilitate the task for our service provider for a rapid adoption and implementation of this new approach by defining the contracts to respect for both parties: VEOLIA and his service provider.

8 Conclusion

In this paper, we proposed an efficient solution for extending the technical capabilities and performance of a platform, based on the OpenStack free solution for the creation of two flexible clouds. The first private cloud contains our current platforms and the second cloud Hosted by our service provider will serve as a platform for reinforcement when necessary in the event of internal overhead, this approach will allow us to minimize future costs and budgets, which will be limited to a resource allocation according to the timing of consumption in place Purchase and day-to-day management.

The idea is to apply this approach to all of the company's subsidiaries worldwide, a private cloud for each subsidiary and then consolidate them into a single global private cloud that will be supported by our service provider's public cloud When the configurable thresholds are exceeded.

ACKNOWLEDGMENT

This work was carried out with the support of Our home affiliation, for the implementation of the "VALVEOL - Cloud" project.

We thank all members of the project team who contributed to the success of this proposal.

REFERENCES

[1] Haifeng Li, Huachun Zhou, Hongke Zhang, Bohao Feng, "EmuStack: An OpenStack-Based DTN Network Emulation Platform", 2016 International Conference on Networking and Network Applications (NaNA), vol. 00, no. , pp. 387-392, 2016, doi:10.1109/NaNA.2016.24

[2] Dinkar Sitaram, Sudheendra Harwalkar, K.V. Suchith Kumar, "Standards Based Integration of Intercloud for Federation with OpenStack", 2016 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), vol. 00, no. , pp. 113-118, 2016, doi:10.1109/CCEM.2016.028

[3] Pooya Musavi, Bram Adams, Foutse Khomh, "Experience Report: An Empirical Study of API Failures in OpenStack Cloud Environments", 2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE), vol. 00, no. , pp. 424-434, 2016, doi:10.1109/ISSRE.2016.42

[4] M. Abhishek Sharma, Monica O. Joshi, "Openstack Ceilometer Data Analytics & Predictions", 2016 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), vol. 00, no. , pp. 182-183, 2016, doi:10.1109/CCEM.2016.045

[5] Ali Babar, Ben Ramsey, "Tutorial: Building Secure and Scalable Private Cloud Infrastructure with Open Stack", 2015 IEEE 19th International Enterprise Distributed Object Computing Workshop (EDOCW), vol. 00, no. , pp. 166, 2015, doi:10.1109/EDOCW.2015.23

[6] Dario Bruneo, Francesco Longo, Giovanni Merlino, Nicola Peditto, Carmelo Romeo, Fabio Verboso, Antonio Puliafito, "A Modular Approach to Collaborative Development in an OpenStack Testbed", 2015 IEEE Fourth Symposium on Network Cloud Computing and Applications, vol. 00, no. , pp. 7-14, 2015, doi:10.1109/NCCA.2015.12

[7] Vojtech Cima, Bruno Grazioli, SeÃ¡n Murphy, Thomas Michael Bohnert, "Adding energy efficiency to Openstack", 2015 Sustainable Internet and ICT for Sustainability (SustainIT), vol. 00, no. , pp. 1-8, 2015, doi:10.1109/SustainIT.2015.7101358.

[8] D. Thomas Erl, D. Zaigham Mahmood, D. Ricardo Puttini, “Cloud Computing: Concepts, Technology & Architecture” September 2015.

[9] Yun Zhang, Farhan Patwa, Ravi Sandhu, Bo Tang, "Hierarchical Secure Information and Resource Sharing in OpenStack Community Cloud", 2015 IEEE International Conference on Information Reuse and Integration (IRI), vol. 00, no. , pp. 419-426, 2015, doi:10.1109/IRI.2015.71.

[10] [10] Robayet Nasim, Andreas J. Kassler, "Deploying OpenStack: Virtual Infrastructure or Dedicated Hardware", , vol. 00, no. , pp. 84-89, 2014, doi:10.1109/COMPSACW.2014.18.

[11] David Bernstein, "Cloud Foundry Aims to Become the OpenStack of PaaS", IEEE Transaction on Cloud Computing, vol. 1, no. , pp. 57-60, July 2014, doi:10.1109/MCC.2014.32.

[12] Juan Angel Lorenzo del Castillo, Kate Mallichan, Yahya Al-Hazmi, "OpenStack Federation in Experimentation Multi-cloud Testbeds", , vol. 02, no. , pp. 51-56, 2013, doi:10.1109/CloudCom.2013.103.

[13] Yang Wang, Xiaomin Wang, Jianyong Chen, "On-Demand Security Architecture for Cloud Computing",

IEEE Computer, vol. 45, no. , pp. 73-78, July 2012, doi:10.1109/MC.2012.120.

[14] V.KRISHNA REDDY1 , Dr. L.S.S.REDDY, "Security Architecture of Cloud Computing",International Journal of Engineering Science and Technology (IJEST), ISSN : 0975-5462 Vol. 3 No. 9 September 2011 pp.7149-7155.

[15] A Beloglazov, J Abawajy, R Buyya - Future generation computer systems, Future Generation Computer Systems, Elsevier, Volume 28, Issue 5, May 2012, Pages 755–768.

[16] Ian Foster, Rubén S. Montero, Borja Sotomayor, Ignacio M. Llorente, "Virtual Infrastructure Management in Private and Hybrid Clouds", IEEE Internet Computing, vol. 13, no. , pp. 14-22, September/October 2009, doi:10.1109/MIC.2009.119.

[17] Wen-Hwa Liao, Shuo-Chun Su, "A Dynamic VPN Architecture for Private Cloud Computing", 2011 IEEE 4th International Conference on Utility and Cloud Computing (UCC 2011), vol. 00, no. , pp. 409-414, 2011, doi:10.1109/UCC.2011.68.

[18] Sumit Goyal, "Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review", I.J. Computer Network and Information Security, 2014, 3, 20-29.

[19] R. Palwe, G. Kulkarni and A. Dongare, —A new approach to hybrid cloud,‖ International Journal of Computer Science and Engineering. 2(1), 1-6, 2012.

[20] L. Cao, X. Liu, M. Liu and K. Han, —Processbased Security Detection Approach for Virtual Machines on Private Cloud Platforms,‖ Journal of Networks, 8(6), 1380-1386, 2013.

[21] T.S. Soares, M.A.R. Dantas, D. Macedo and M.A. Bauer, —A data management in a private cloud storage environment utilizing high performance distributed file systems,‖ In: Proceedings of the IEEE 22nd International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises,

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Optimizing Hadoop for Small File Management

O. Achandair, M. Elmahouti, S. Khoulji, M.L. Kerkeb,
Information System Engineering Resarch Group
Abdelmalek Essaadi University, National School of Applied Sciences, Tetouan, Morocco.
o.achandair@gmail.com; mouradelmahouti@gmail.com; khouljisamira@gmail.com;
kerkebml@gmail.com

ABSTRACT

HDFS is one of the most used distributed file systems, that offer a high availability and scalability on low-cost hardware. HDFS is delivered as the storage component of Hadoop framework. Coupled with map reduce, which is the processing component, HDFS and MapReduce become the de facto platform for managing big data nowadays. However, HDFS was designed to handle specifically a huge number of large files, while when it comes to a large number of small files, Hadoop deployments may be not efficient. In this paper, we proposed a new strategy to manage small files. Our approach consists of two principal phases. The first phase is about consolidating more than only one client's small files input, and store the inputs continuously in the first allocated block, in a SequenceFile format, and so on into the next blocks. That way we avoid multiple block allocations for different streams, to reduce calls for available blocks and to reduce the metadata memory on the NameNode. This is because groups of small files packaged in a SequenceFile on the same block will require one entry instead of one for each small file. The second phase consists of analyzing attributes of stored small files to distribute them in such a way that the most called files will be referenced by an additional index as a MapFile format to reduce the read throughput during random access.

Keywords - Cloud Hadoop, HDFS, Small Files, SequenceFile, MapFile

1 Introduction

The exponential growth of data, generated continuously in different formats and from different sources, has lead traditional architecture and infrastructures to face many limitations. In the last decade, new technologies based on the cloud model, offered to many organizations the possibility to store and analyze their data in an efficient way and a timely manner, which help them uncover patterns, get insight and provide better services. Hadoop, is an open source framework that offers a distributed storage layer, HDFS [1], tightly coupled with a distributed processing engine,MapReduce [2]. Hadoop allowsthe partitioning of data and computation across clusters of thousands of machines, in such a way, that each machine compute its local or neighbor's data. Hadoop cluster offers high scalability simply by adding commodity servers, as each server can hold more data and process more tasks. The total computation capacity andstorage capacity can be expanded, in a transparent and an efficient manner. Hadoop guarantee a high fault tolerance and high availability. In fact, in the storage layer, HDFS replicate blocks between nodes, so whenever a machine of the cluster goes down, the data can still be accessed from others. Also, HDFS

DOI: 10.14738/tmlai.54.3214
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3214

maintain the replication factor by replicating blocks to other available machines in case of failures. Moreover, in the processing layer, Hadoop can keep track of all the tasks, and restart them on other available machines whenever a host-failure occur during tasks processing. Hadoop is an Apache top-level project, built in modular approach, thatincludes multiple components and subprojects. The components of Hadoop ecosystem are classified as shown in "Fig. 1"

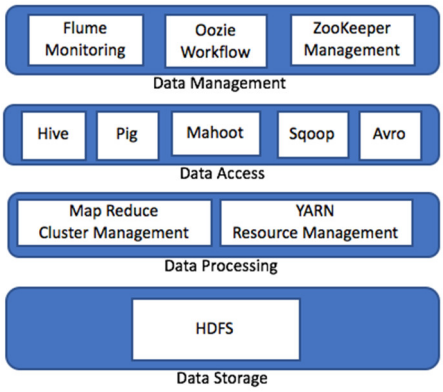


Figure. 1: Hadoop Data Layers Stack

The Data Storage layer, consist of HDFS as the main component that provide the physical access for read and write to the cluster. The Data Processing layer consists of MapReduce as the framework that enables users to write applications that can be processed in parallel according the MapReduce programming model, and YARN, as the resource management component. YARN keeps tracks of resources on the cluster and orchestrates all the assigned tasks. The Data Access layer, consist of the infrastructure that offer tools to manipulate and analyze data through scripting, modeling, querying. The Management layer, consists of the end user layer, it addresses data governance, integration and monitoring components.Hadoop has contributed hugely to handle and manage big data, itsstorage layer was designed to store and process large big files, which are in gigabytes or terabytes, but when it comes to a large number of small files, the performance of the cluster may decrease dramatically. In this research, we addressed the small file problem through an additional middleware called Small File Analyzer server (SFA). The SFA component interacts directly with the data storage and the data processing layers.

The rest of this paper is divided into the following, section 2 gives more details about the data access layer and the data processing layer. Section 3 lists the existing solutions in the related literature. Section 4 present the proposed approach. Section 5 is allocated for our experimental works and results. Finally, Section 5 for conclusion and expectation.

2 Backround

2.1 Hadoop Storage:

HDFS, The Hadoop distributed file system provides high reliability, scalability and fault tolerance. It's designed to be deployed on big clusters of commodity hardware. It's based on a master-slave architecture, the NameNode as a master and the DataNodes as slaves. The NameNode is responsible for managing the file system namespace, it keeps tracks of files during creation, deletion, replication [3] and manages all the related metadata [4] in the server memory. The NameNode splits files into blocks and sends the writes requests to be performed locally by DataNodes. To ensure a fault-tolerance system, blocks replicas are

pipelined across a list of DataNodes. This architecture as shown in "Fig.2", with only one single NameNode simplifies the HDFS model, but it can cause memory overhead and reduces file access efficiency when dealing with a high rate of small files.

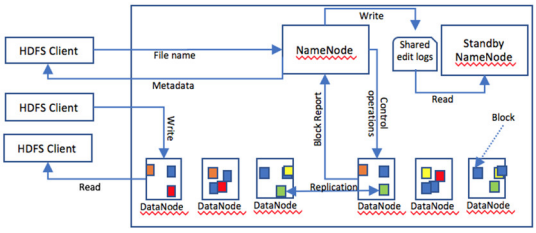


Figure. 2: HDFS Architecture

2.2 Hadoop Processing:

In the current version of Hadoop, Google re-architected the processing engine to be more suitable for most of big data applications needs. The major improvement of Hadoop was the introduction of a resource management module, called YARN, independently of the processing layer. This brought significant performance improvements, offered the ability to support additional processing models, and provided a more flexible execution engine. Because of its independency architecture, existing MapReduce applications can run on YARN infrastructure without any changes.

The MapReduce program execution on YARN can be described as follows:

(1) A user defines an application by submitting its configuration to the application manager
(2) The resource manager allocates a container for the application manager
(3) Resource manager submits the request to the concerned node manager
(4) The Node manager launches the application manager container
(5) The application manager gets updated continuously by the node manager nodes, it monitors the progress of tasks

When all the tasks are done, the application manager unregisters from the resource manager, like so, the container can be allocated again, See "Figure. 3".

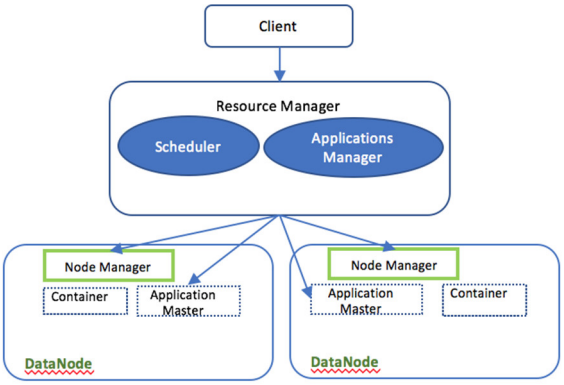


Figure. 3: YARN – Yet Another Resource Manager

To deal with the small file problem, numerous researchers have proposed different approaches. Some of those efforts have been adopted by Hadoop and are available for use natively, more precisely, Hadoop Archives (HAR Files) and SequenceFile.[5]

3 Small File Problem

In HDFS, a file is called small when its size is smaller than the size of the HDFS block. Each file or block is an object in the namespace, each object occupies around 150 bytes in the NameNode memory as the related metadata. Consider a file of 1Gb to be stored in HDFS, with a default block size of 64 Mb and a default replication factor of 3, then we will need 16x3 blocks on the DataNodes and 16x3x150=2400 bytes in the NameNode memory. Instead, if we consider 1000 files of 1Mb, and assume that each file will be stored independently on a block.Then the physical storage on DataNodes will remain the same as the 1Gb file, but we will need 600 000bytes in the NameNode memory. This is because there will be one entry per block which is 1000x3 (number of blocks x replication factor) and one entry per file which is 1000. Each block or file entry will take about 150bytes. As a result, for the same physical storage250 times additional memory space will be required, compared to the previous example. As a matter of fact, there is many fieldsthat produce tremendous numbers of small files continuouslysuch as analysis for multimedia data mining [6], astronomy [7], meteorology [8], signal recognition [9], climatology [10,11], energy, and E-learning [12] where numbers of small files are in the ranges of millions to billions. For instance, Facebook has stored more than 260billion images [13]. Inbiology, the human genome generates up to 30 million files averaging 190KB [14]. Massive numbers of small files can decrease dramatically the NameNode performance, as for each file access, the HDFS client needs to retrieve the metadata from the NameNode. Therefore, frequent calls for and frequent access to metadata reduce the latency during read and write throughput.

In terms of Hadoop processing, the time needed to process too many small files can be hundreds of times slower than processing one single large file that has the same total size. In fact, under a default configuration, Hadoop creates a mapper for each file. Like so, we will have a great number of mappers, that are costly resources.

4 Related Work

To deal with the small file problem, numerous researchers have proposed different approaches. Some of those efforts have been adopted by Hadoop and are available for use natively, more precisely, Hadoop Archives (HAR Files) and SequenceFile.

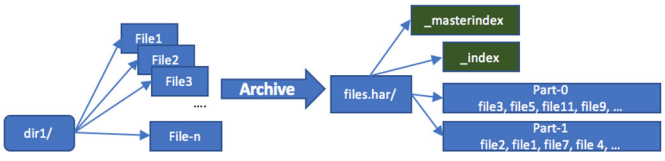


Figure. 3: HAR File Layout

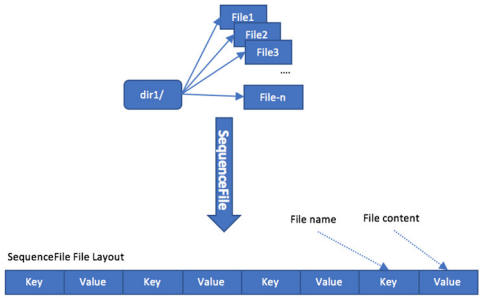


Figure. 4: SequenceFile File Layout

Hadoop Archive packs small files into a large file, so that we can access original files transparently, see "Fig. 3". This technique allows more storage efficiency, as only metadata of the archive is recorded in the namespaceof the NameNode, but it doesn't resolve other constraints in terms of reading performance.Also, the archive cannot be appended while adding more small files. The SequenceFile technique is to merge a group of small files ina flat file, as key-value pairs, while key is the related file metadata and value is the related content, see "Fig. 4".

Unlike the HAR files, the SequenceFile supports compression, and they are more suitable for MapReduce tasks as they are splittable [15], so mappers can operate on chunks independently.However, converting into a SequenceFile can be a time-consuming task, and it has a poor performance during random read access.

To improve the metadata management, G. Mackey et al. [16] merged small files into a single larger file, using the HAR through a MapReduce task. The small files are referenced with an added index layer (Master index, Index) delivered with the archive, to retain the separation and to keep the original structure of files.

C. Vorapongkitipunet al. [17] proposed an improved approach of the HAR technique, by introducing a single index instead of the two-level indexes. Their new indexing mechanism aims to improve the metadata management as well as the performance during file access without changing the implemented HDFS architecture.

Patel A et al. [18] proposed to combine files using the SequenceFile method. Their approach reduces memory consumption on the NameNode, but it didn't show how much the read and write performances are impacted.

Y. Zhang et al. [19] proposed merging related small files according to WebGIS application, which improved the storage effeciency and HDFS metadata management, however the results are limited by the scene.

D. Dev et al. [20] proposed a modification of the existing HAR. They used a hashing concept based on the sha256 as a key. This can improve the reliability and the scalability of the metadata management, also the reading access time is greatly reduced, but it takes more time to create the NHAR archives compared to the HAR mechanism.

P. Gohil et al. [21] proposed a scheme for combining small file, merging, prefetchingthe related small files which improves the storage and access efficiency of small files, but does not give an appropriate solution for independent small files.

5 The Proposed Approach For Small File Management

Currently, it's well known that small files can decrease dramatically the performance of Hadoop clusters. Previous solutions workaround this problem by packaging small filesin different formats. Those formats are saved transparently in HDFS as they can be divided into blocks with no specific constraints. Though the performance of MapReduce jobs can be greatly improved based on the way those small files are packaged, none of the adopted mechanisms take in consideration how to organize those small files during the merging phase. The core idea behind our approach is to store files when a client starts a stream that containssmall files, combined if relevant, with other client streams into a large file within the same block, and organize them later in an efficient way that we can prefetch the most probable called files first. This was achieved by using a Small File Analyzer, see “Fig. 5”.

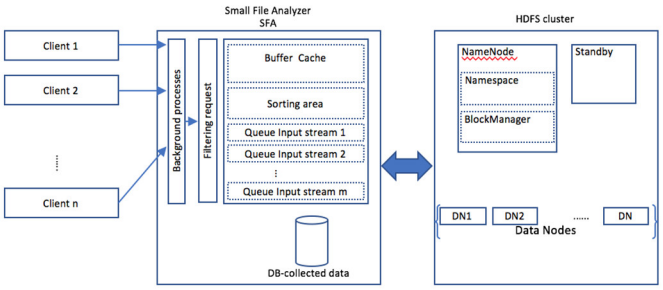


Figure. 5: SFA Architecture

The SFA operations consists of two phases, the first one consists of combining similar small files and store them on one block.

The second one consists of analyzing how small files are used, then put adequate groups in a MapFile. A MapFile is another format of packaged files offered by Hadoop, that consists of an indexed SequenceFile. See “Fig 6”.

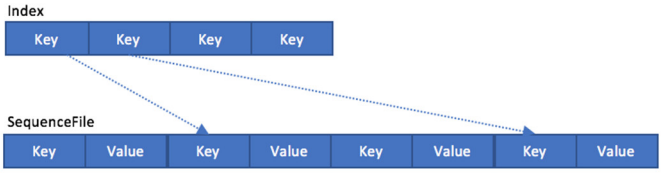


Figure. 6: MapFile File Layout

This second phase can be triggered or called manually as analyzing small files dependson how the cluster is using them, getting the most called small files depends on how many jobs are using them, also on the fact that the files are called together sequentially or extracted in a very few groups.

To improve the SFA analysis, we import records from the FSimageof the NameNode, which contain a complete state of the HDFS state, then we aggregate data to store it in a reference handled by the database of the SFA, see “Fig. 7”. All the clients’ jobs are forwarded from the SFA server first, in such a way, we keep track of more information to use in the SFA analysis.

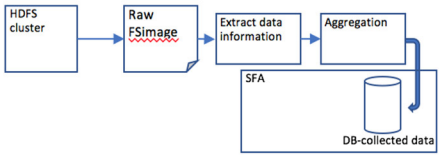


Figure. 7: MapFile File Layout

5.1 Presentation of The Two Phases Algorithms

Algorithm phase1: Filtering inputs & storing small files

Input: Client dataset

Output: SequenceFile of combined small on the same blockid

Step1: Get stream characteristics

Step2: Initialize combined queues

Step3: Request blockid from NameNode and lock it in the SFA blocklist.

Step4: Merge maximum small files into a SequenceFile to the locked blockid

Step5: Close current output stream

Free the block from the block list.

If the current block is full, NameNode will allocate a new block, and the full one is deleted from the SFA list.

To use the block capacity efficiently, SFA chains small file queues when different clients are requesting storage in HDFS at the same time. Based on the grouped small files in this step2, SFA will request a list of blocks. Each time a blockid is full, SFA deletes it from its blocklist. Like this SFA will write the next coming request to the first blockid in its list without requesting each time a new block from the NameNode.

Algorithm phase2: Analyzing Small Files

Input:SFA inventory - state j

Output: MapFile of Hot files - SFA inventory state j+1

Step1: Download and Interpret FSimage using oiv interpreter

Step2: Splitting FSimage rows and initialize SFA reference

Step3: Aggregate records

Step4: Get Top called sequence

Step5: Get Top called files per sequence

Step6: Define group of hot small files

Step7: Retrieve the hot files from original sequence and merge them in a MapFile

---

Once the hot files are listed, we can get the keys and values from their original SequenceFile, and create theMapFiles. SFA can schedule the migration in off-hours. This operation consists of three parts as follows:

```
// get the keys and values of the hot file list
SequenceFile.Reader reader = null;
Class<? extends Writable>keyClass= null;
Class<? extends Writable>valueClass= null;
try{reader= new SequenceFile.Reader(fs, sequenceFile, getConf());
keyClass= (Class<? extends Writable>) reader.getKeyClass();
valueClass= (Class<? extends Writable>) reader.getValueClass();
} catch (IOExceptionioe) {
MainUtils.exitWithStackTraceAndError(
"Failed to open SequenceFile to determine key/value classes: " + input, ioe);
} finally {
if (reader != null) {
try {
reader.close();
}}}

// move the SequenceFile to the new map file , rename it within the output location
```

```
try {

fs.rename(sequenceFile, mapData);

} catch (IOExceptionioe) {

MainUtils.exitWithStackTraceAndError(

"Failed to move SequenceFile to data file in MapFile directory: input=" + input + ", output="

+ output, ioe);

}

// create the index file for the MapFile

try {

MapFile.fix(fs, mapFile, keyClass, valueClass, false, getConf());

} catch (Exception e) {

MainUtils.exitWithStackTraceAndError("Failed to create MapFile index: " + output, e);

}

return 0; }
```

6 Performance Evalation

The proposed method in this paper is compared with the original HDFS about the usage of NameNode memory, and the performance of MapReduce jobs, during sequential and selective file access. We did a simulation on the Hadoop-2.4.0, our cluster consists of one NameNode 3.10 GHz clock speed, 8GB of RAM and a gigabit Ethernet NIC, and four DataNodes. All the nodes offer 500GB Hard Disk, and they are deployed on Ubuntu 14.04. The replication factor is kept as the default value 3 and the block size of HDFS is chosen as 64Mb. The experimental datasets are basically standard auto-generated.

Comparison of the NameNode Memory Usage

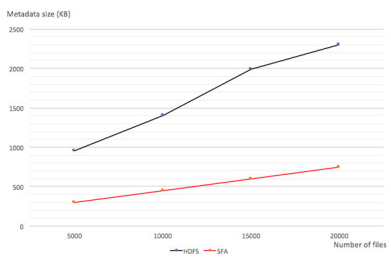


Figure. 8: NameNode memory usage

According to “Fig. 8”, the NameNode memory usage of the original HDFS is biggest due to metadata entries for each small file and the inefficiency of block allocation. When we store small files through SFA, The NameNode memory consumption is too low due to file merging in Sequence File. But this doesn’t show if our block allocation strategy is efficient or not. This will be tested in the next steps of this research, as we are introducing more factors to combine streams in the locked block.

6.1 Comparison of MapReduce jobs performance

In the following, we performed MapReduce jobs on four datasets. The results on HDFS refers to the measurementof time requiredof MapReduce job on SequenceFiles without retrieving the hot files. We ignored storing files without merging them in SequenceFile, as the low latency during MapReduce jobs has been already proved in the previous studies. In fact, this will lead to the creation of huge number of mappers and finally hanging the whole cluster. The results on SFA2 and SFA6 refer to themeasurementof the time required of MapReduce job after the second call and the sixth job iteration. Each measurement is performed after reorganizing small filesas suggested from the SFA

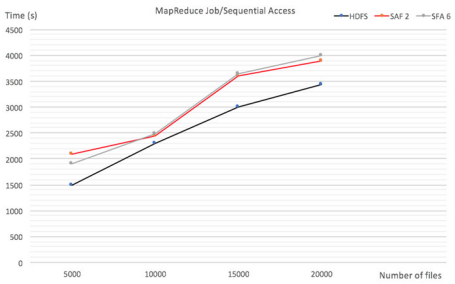


Figure. 9 MapReduce Performance/Sequential Access

According to “Fig. 9”, even after many iteration of MapReduce jobs, reorganizing small files didn’t improve the MapReduce performance, this is because of accessing indexes is not a necessity in such situations.

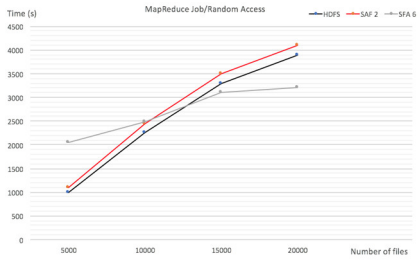


Figure. 10 MapReduce Performance on selective files

According to “Fig. 10”, after six iteration, we observed that the performance of MapReduce job is slightly improved when the dataset gets bigger. The specific MapReduce jobperformed about 13minutes faster than the first execution. When the number of small files is too high, retrieving more adequate hot files to be grouped in a MapFile was a solution. Reading hot files in such a situationthrough an index improves the performance of that specific MapReduce job.

In the next step of our research, we will adjust the SFA by introducing more factors to combine related streams efficiently in the allocated blocks. Also, we will introduce the concept of cycles, as even if the

MapReduce jobs tends to call data independently, analyzing file calls during frequents periods of time can reveal new correlations.

7 Conclusion

Our approach provides a new allocation strategy for blocks when storing massive amounts of small files, it also addresses the aspect of analyzing the distribution of small files in SequenceFile format. This approach of classification of metadata based on number of calls can also be extended to include other factors such as owners, size, and age of datasets that are supported in the initial design of our SFA sever. Different formats are now supported in Hadoop to solve the small files problem, but there is a lack of standardization, as most of the solutions remain useful in specific environmentsbut not in others. Offering a system to analyze different aspects of the small files problem can help organizations to understand better the real factors that control the impact of their datasets.

REFERENCES

[1] Official Hadoop website, http://www.hadoop.apache.org.

[2] J. Dörre, S. Apel, and C. Lengauer, “Modeling and optimizing MapReduce programs,” Concurrency and Computation: Practice and Experience, vol. 27, no. 7, pp.1734-1766, 2015.

[3] D. T. Nukarapu, B. Tang, L. Wang, and S. Lu, “Data replication in data intensive scientific applications with performance guarantee,” IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 8, pp. 1299–1306, 2011.

[4] Y. Gao and S. Zheng, "A Metadata Access Strategy of Learning Resources Based on HDFS," in proceeding International Conference on Image Analysis and Signal Processing (IASP), pp. 620—622, 2011.

[5] T. White. Hadoop: The Definitive Guide, 4th Edition O’Reilly, 2015.

[6] B. White, T. Yeh, J. Lin, and L. Davis, “Web-scale computer vision using mapreduce for multimedia data mining,” in Proceedings of the Tenth International Workshop on Multimedia Data Mining. ACM, 2010, p. 9.

[7] K. Wiley, A. Connolly, J. Gardner, S. Krughoff, M. Balazinska, B. Howe, Y. Kwon, and Y. Bu, “Astronomy in the cloud: using mapreduce for image co-addition,” Astronomy, vol. 123, no. 901, pp. 366–380, 2011.

[8] W. Fang, V. Sheng, X. Wen, and W. Pan, “Meteorological data analysis using mapreduce,” The Scientific World Journal, vol. 2014, 2014.

[9] F. Wang and M. Liao, “A map-reduce based fast speaker recognition,” in Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on. IEEE, 2013, pp. 1–5.

[10] K. P. Ajay, K. C. Gouda, H. R. Nagesh, “A Study for Handelling of High-Performance Climate Data using Hadoop, Proceedings of the International Conference, pp: 197-202, April 2015.

[11] Q. Duffy, J. L. Schnase, J. H. Thompson, S. M. Freeman, and T. L. Clune, “Preliminary evaluation of mapreduce for high-performance climate data analysis,” 2012.

[12] C. Shen, W. Lu, J. Wu, and B. Wei, “A digital library architecture supporting massive small files and efficient replica maintenance,” in Proceedings of the 10th Annual Joint Conference on Digital Libraries (JCDL '10), pp. 391–392, June 2010

[13] A. Thusoo, Z. Shao, S. Anthony, D. Borthakur, N. Jain, J. Sen Sarma, R. Murthy, and H. Liu, “Data warehousing and analytics infrastructure at facebook,” in Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010, pp. 1013–1020.

[14] J. K. Bonfield and R. Staden, “ZTR: A new format for DNA sequence trace data”, Bioinformatics, vol. 18, no. 1, (2002), pp. 3–10.

[15] J. Xie, S. Yin, et al. “Improving MapReduce performance through data placement in heterogeneous Hadoop clusters”, In 2010 IEEE International Symposium on Parallel & Distributed

[16] G. Mackey; S. Sehrish; J. Wang. Improving metadata management for small files in HDFS. IEEE International Conference on Cluster Computing and Workshops (CLUSTR). 2009. pp.1-4.

[17] C. Vorapongkitipun; N. Nupairoj. Improving performance of small-file accessing in Hadoop. IEEE International Conference on Computer Science and Software Engineering (JCSSE). 2014. pp.200-205.

[18] Patel A, Mehta M A. A novel approach for efficient handling of small files in HDFS, 2015 IEEE International Advance Computing Conference (IACC), pp. 1258-1262.

[19] Y. Zhang; D. Liu. Improving the Efficiency of Storing for Small Files in HDFS. International Conference on Computer Science & Service System (CSSS). 2012. pp.2239-2242

[20] D. Dev; R. Patgiri. HAR+: Archive and metadata distribution! Why not both?. IEEE International Conference on Computer Communication and Informatics (ICCCI). 2015. pp.1-6.

[21] P. Gohil; B. Panchal; J. S. Dhobi. A novel approach to improve the performance of Hadoop in handling of small files. International Conference on Electrical, Computer and Communication Technologies (ICECCT). 2015. pp.1-5.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

MEC towards 5G: A Survey of Concepts, use Cases, Location Tradeoffs

Belghol Hibat Allah and Idrissi Abdellah
IPSS Team, Computer Science Department, Faculty of Sciences
Mohammed V University in Rabat
hibatbelghol@gmail.com; idrissi@fsr.ac.ma

ABSTRACT

In recent years, there has been a new trend to push everything to a centralized cloud, triggered by virtualization and pushed by the need to reduce costs and increase suppleness. In the process, mobile operators and industry players forgot how prominent the location of the functionality is to performance, optimal use of network resources and user experience. As they progressively grasp the influence of location in wireless networks and specifically in virtualized networks they start to look for efficient ways to deploy network taking in consideration those metrics. On the one hand Distributed Cloud RAN (DC-RAN) which consists in deploying Base Band Units (BBUs) in distributed way instead of pooling the units at centralized data center and Multiple-access Edge Computing (MEC), Previously known as Mobile Edge Computing which consists in mixing the IT & Telco domains and bringing their capabilities within the close proximity of mobile subscribers to better serve them are gaining acceptance in mobile networks. On the other hand, Software Defined Networking (SDN) and Network Function Virtualization (NFV), two promising concepts are expected to take mobile networks to a high level of agility. While SDN is based on the separation of control and data planes, NFV separates software from hardware enabling flexible network deployment and dynamic operation. Virtualization, originally used as the support for shifting to the centralized cloud, is even more basic in permitting hybrid models because it offers service providers the chance to choose a location, hardware, and software separately to optimize end-to-end network performance and Quality of Experience (QoE). In this perspective, NFV, SDN, DC-RAN, and MEC in distinct but in complementary ways man up the necessity to put processing and storage where it's suitable to preserve a healthful harmony between what lingers centralized and what have to be distributed to the edge based on parameters such as applications, traffic type, and network conditions. In this paper, we will present MEC use cases that have gained an attraction to date and we will shed the light on the importance of the edge location and criteria to take in consideration while deploying MEC within network along with the MEC location tradeoffs.

Keywords: MEC, NFV, SDN, DC-RAN, 5G, EDGE.

1 Introduction

The transition from LTE to 5G will be the deepest revolution in the wireless industry since the evolution from analog to digital [1]. There are many reasons for this, primarily because new uses for wireless technology are creating demand for new approaches to connectivity, bandwidth, and network

DOI: 10.14738/tmlai.54.3215
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3215

architecture; 5G is expected to fulfill a very wide range of requirements such as higher bandwidth, ultra-dense network, very low latency and huge capacity. This evolution will bring about several new ways of designing networks so that the promise of always-on, high-bandwidth, low latency and massive networks can become reality. According to the last Cisco forecasts, the overall mobile data traffic especially the video one in its multiple forms; e.g, downloaded, streamed, uploaded, interactive is expected to know a phenomenal growth by 2021, a seven-fold increase over 2016. This tremendous rise of traffic is driven mainly by five major trends: evolution toward Smarter Mobile Devices, defining Cell Network Advances - the evolution of 4G and the coming introduction of 5G, mass mobile Internet of Everything (IoE) adoption, versatility of Wi-Fi and expansion of its use and coverage and identification of new mobile applications and requirements [2]. However, due to resource constraints (processing power, battery lifetime, and storage capabilities), mobile devices are unable to perform effectively and cater for this myriad of bandwidth-intensive applications with stringent requirements, which calls for Cloud Computing (CC) involvement. Thus, the traditional CC centralized model (processing, storage...) does not align with the 5G requirements and may lead to some limitations such as high latency, limited bandwidth, and scalability problems [3].

IT and Telecom Networking domains are converging bringing with them new opportunities and capabilities that can be deployed into the network. A key revolution has been the capacity to run IT servers at network edge nearby the mobile subscriber applying cloud computing concept, we define this as MEC. This is one of the most relevant initiatives proposed during the past few years to address the above challenges. MEC speeds up content, services, and applications, augmenting reactivity from the edge. Laying by this the ground for e-health, self-driving cars, augmented reality, intelligent houses and cities, enhanced mobile broadband and many other applications. The mobile subscriber's experience can be enhanced through effective network and service operations, based on an understanding of the radio and network conditions [4]. In today's networks, the dominant approach is to boost performance, given the accessible financial resources. This generally means enhancing network capacity using a pressure approach by deploying the recent technological features that increase throughput. But far less effort is put into optimizing the use of the capacity available. There are many ways to optimize network resource utilization and, under severe pressure to improve performance without increasing costs, mobile operators have started to take steps toward achieving this optimization goal. At the RAN level, two approaches have to be considered: DC-RAN deploys BBUs in distributed manner nearby the Remote Radio Unit (RRU) instead of pooling the units at centralized data center while MEC is in general geared to balance the processing, storage and control in the network by moving processing, storage and control to the edge of mobile network, but the edge location and its prominence remains blur which leads us to explore the different possible locations and assess their impact on the global architecture. For the core network, SDN-NFV integration could bring huge benefits enabling a flexible, agile and scalable network. The remainder of this paper is structured as follows: An overview of MEC is given in Section II. Section III, lists MEC use cases gaining attraction then the importance of the edge location, as well as location tradeoffs are examined in depth. Finally, Section IV concludes the paper.

2 Overview of Multiple-Access Edge Computing (MEC)

Among the different actions taken to realize edge computing and, more specifically, mobile edge computing, European Telecommunications Standards Institute (ETSI) MEC occupies a central role because it provides a structure to move processing, storage, and control to the edge. MEC was conceived to

respond to mobile operators need to shift the processing and storage of some services and applications to the brink, and to improve mobile network performance and resource usage in real time. MEC standardization work began in 2014 at ETSI, with a seminal white paper authored by the founders of the MEC ISG Huawei, IBM, Intel, Nokia, NTT DOCOMO, and Vodafone [5]. This white paper discusses the business, consumer and technical value/benefits offered by this technology, the enablers, the requirements and challenges for MEC as well as the objectives of the MEC initiative. Additionally, it presents the high-level architectural blueprint of MEC as well as a number of use cases where MEC seems to have a prominent impact. Content caching, application computation offloading and connected vehicles are examples of uses cases among other existing. The list of active participants has since expanded to cover more vendors and operators. The focal point was first the mobile networks, but it has extended to fixed networks, to mirror the close amalgamation of mobile and fixed networks, which often host the same services and applications. The MEC acronym not any more assigned to Mobile Edge Computing and rather refers to Multiple-access Edge Computing. Albeit the change was catalyzed by ETSIs internal process requirements, it was a welcome change as it expands the reach and potential of MEC in today's rapidly evolving wireless networks, which consist of a various set of access technologies and spectrum regulatory structures. The extension to non-cellular technologies means that Wi-Fi is now included within MECs scope. This is a welcome addition that mirrors the fact Wi-Fi in the majority of markets stands for the bulk of traffic to mobile devices. The ETSI MEC standardization work creates the foundation for edge computing deployments in mobile and, progressively, fixed networks. During the first term (2015-2017), ETSI ISG finished the basic work, delivered the elemental specifications, and emboldened the creation of the ecosystem. More work is needed during the second terms (2017-2018) not only to extend beyond mobile networks but also to bolster the connections with other edge computing initiatives while evading the fragmentation of exertions. Beyond standardization and industry collaborative initiatives, there is a need to investigate various business and deployment models and review the role that stakeholders will have in the development chain [6]. MEC and other promising edge computing initiatives such as Fog Computing, Cloudlets address the need to place processing and storage at the network edge in close proximity to the final user. In this context, a high-level overview was proposed in [7] to show that those three approaches include partially overlapping concepts and are also complementary. In many ways, edge computing concept is nothing new. There has been edge computing solutions all the time to respond to a specific needs and optimization issues in mobile networks. What differs today is that network virtualization gives a framework to extend edge computing capacities, adding adaptability, trustworthiness, and elasticity. This will permit to operators to garner profits in terms of improved QoE and resource usage. While MEC and C-RAN are not straight linked to each other, the C-RAN function, in theory, could be deployed over any cloud service. However, to make that possible, the front-haul transport between the RRUs at the cell sites and the virtualized BBUs in the centralized data centers would require high throughput and low latency connections. Achieving this kind of transport requires that the centralized BBUs be placed within close proximity to the RRUs. With this perspective, it's conceivable that C-RAN could be implemented at the edge as a way of augmenting the benefits of MEC [8]. And that was the main goal of the work done in [9]. The authors presented in [9] an integration of mobile edge services in scenarios where an aggregation of small cells is able to share certain processing and storage resources in a local perspective. Therefore, rather than centralizing RAN and MEC functions into Big Data center, the proposed solution creates a distributed light data center by coordinating the micro servers associated with different small cells in a cluster. MEC and Fog computing, the two are closely related. Both describe

standards-based architectures where computing, storage, and networking resources are made available in a cloud fashion on host servers located at the network edge. The main difference is in the context in which they are referenced. Fog computing is often mentioned in the context of Internet of Things (IoT), where host servers are typically routers, access points or even computing devices co-located with sensors and actuators. MEC is often mentioned in the context of mobile networks, where host servers are integrated with the mobile network infrastructure, such as base stations or aggregation sites. It is considered as a sub-category of Fog computing. IoT, which is highly associated with the mobile network, will benefit from extending the Software Defined Mobile Edge Computing (SD-MEC) concept to other areas (VANET, WSN, etc.) to enable wide deployment for IoT. In [10], the authors have tried to shed light on these new trends and drawn their vision of the Internet of the future. Based on a virtualized platform, MEC is recognized by the European 5G Infrastructure Public Private Partnership (5G PPP) research body as one of the key emerging technologies for 5G networks together with NFV and SDN [11].

To the best of our knowledge, the integration of MEC, DC-RAN, NFV, SDN and Big Data Analytics in one overall network architecture putting a special focus on the prominence of the choice of the edge location has not been addressed in the existing works. A Cloud-based Wireless Network with virtualized, reconfigurable and smart Wireless Network was proposed by the authors in [12]. However, the architecture didn't include MEC and DC-RAN which are two key factors to boost the network responsiveness and decrease latency to cater for the wide variety of 5G requirements. In essence, this paper will discuss the most relevant MEC uses cases and analyze the importance of the MEC server location and the key parameters to take in consideration in order to ensure the balance of the whole network.

3 MEC Applications Scenarios and Edge Location Tradeoffs

3.1 Applications Scenarios

In this section, we will start by presenting MEC use cases that have attracted attention to date and which are divided into three main categories: Operator and third party services, user-oriented services and network performance and QoE improvements. Then we will discuss the importance of the edge location along with the tradeoffs to be considered.

3.1.1 Intelligent Video Acceleration:

Fig.1 shows a radio analytics application providing the video server with an indication on the throughput estimated to be available at the radio downlink interface. This information can be used to assist TCP congestion control decisions and also to ensure that the application-level coding matches the estimated capacity at the radio downlink and then enables improved video quality and throughput. It uses data about network conditions especially RAN conditions and RAN load to generate advice for content and application providers on how to manage traffic exchanged with the subscriber. When the network has sufficient capacity, the providers can share content at the highest quality available. When the network is capacity constrained or congested, the content and TCP transmission can be adapted to provide subscribers the best experience possible given the real-time availability of network resources [13].

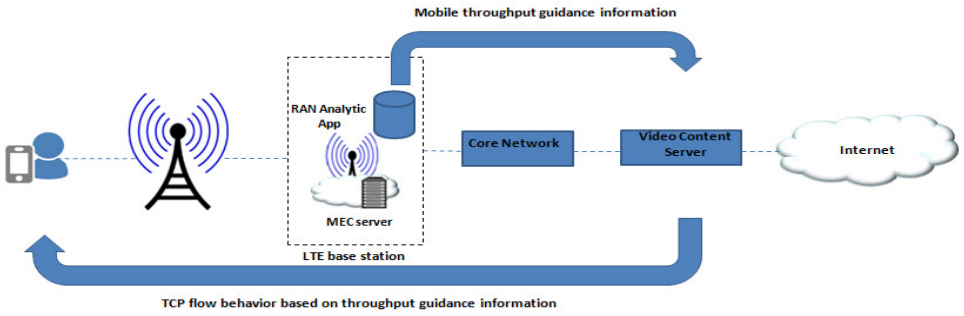


Figure 1. Intelligent Video Acceleration

3.1.2 Local Content Caching:

Fig.2 depicts how subscribers at a given location and time tend to watch a remarkably consistent and narrow set of content so not only traffic is growing, it is often concentrated in specific locations and times that are the backdrop for perfect use case to justify moving processing and storage to the edge. MEC cannot lower the latency in the RAN, but it can reduce the end-to-end latency by storing locally the most popular content consumed in a specific geographical area. Once requested, the content is provided from the local cache. In this case no need to transfer the content always through the core network [13].

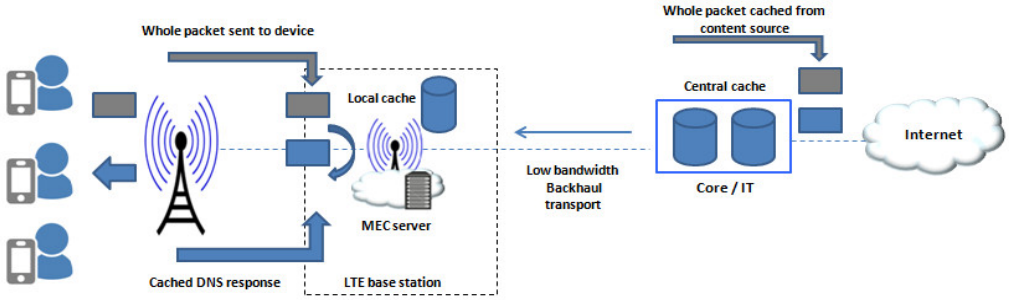


Figure 2. Local Content Caching

3.1.3 Augmented Reality:

Fig.3 refers to Augmented Reality. This MEC application analyses the output from a device's camera and the precise location. Objects viewed on the device camera are overlaid with local augmented reality content. It enables the unique experience of a visitor to a museum or other (indoors or outdoors) points of interest and ensures low latency and high rate of data processing [14].

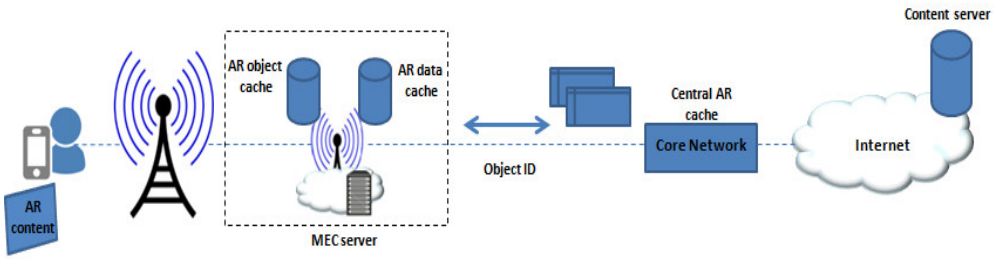


Figure 3. Augmented Reality

Application computation offloading: In Fig.4 the mobile edge host executes compute-intensive functionalities with high performance instead of the mobile devices, improving the user experience and the consumer can use low complexity devices by off-loading compute capacity to the mobile edge host. Therefore, new compute-intensive services like auto-translation and recommendations synthesized from more than one mobile edge application can be delivered in near real time to low complexity devices. In this way, the operator can provide value added services by utilizing application data available at the mobile edge host [13].

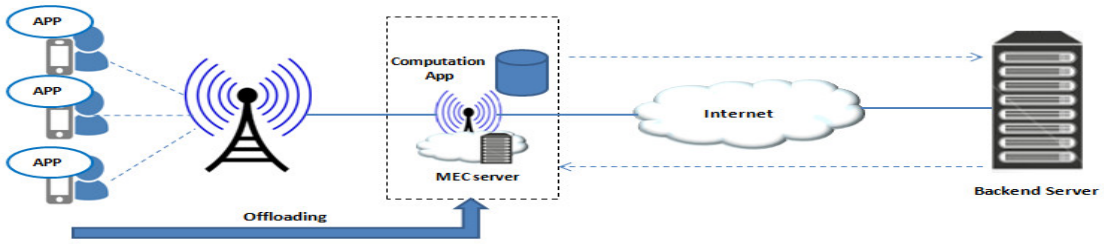


Figure 4. Application computation offloading

3.1.4 Video Analytics:

Fig.5 presents the distributed live video streams analytics application where Events are triggered automatically (e.g. movement, missing objects, crowd) enabling fast detection and action triggering; which is useful and applicable to public safety, smart cities [13].

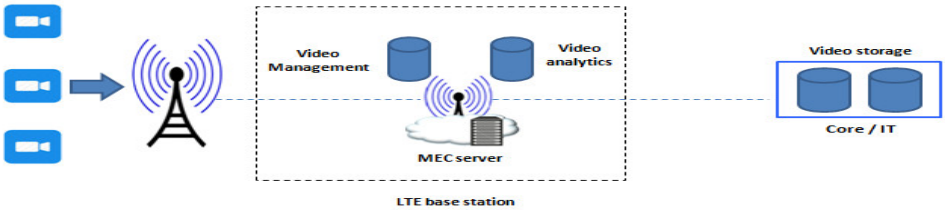


Figure 5. Video Analytics

3.1.5 Location Based Services:

Fig.6 shows how a real-time location of an active device is tracked and provided in a passive way independent of GPS information Mitigating indoor limitation of GPS and helps to locate specific users and understand how the crowd is distributed. It is applicable to Smart City, Geo-Fencing, Retail, and advertising [13].

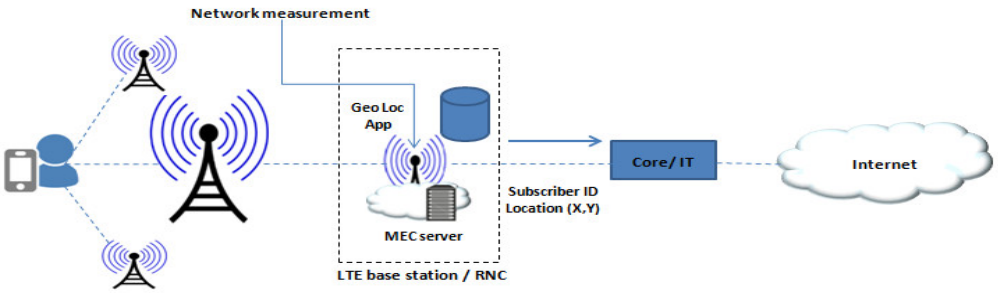


Figure 6. Location Based Services

3.1.6 Connected Vehicles:

Existing cloud services are extended into the highly distributed mobile base station environment, leveraging the existing LTE connectivity. The MEC application presented in fig.7 operates as a roadside unit for vehicle-to-infrastructure (V2I) and road hazards can be recognized and warnings can be sent to nearby cars with extremely low latency which enables a nearby car to receive data in a matter of milliseconds, and the driver to react instantly [14].

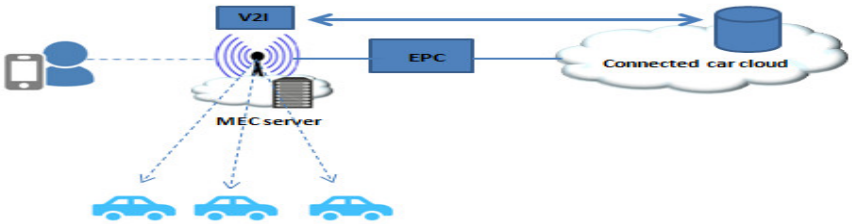


Figure 7. Connected Vehicles

3.2 The prominence of MEC server location

We have talked so far about MEC concept and how shifting processing and storage to the edge of mobile network could bring huge benefits in term of QOE and network performance. However, is not clear at all where the edge could be or, more meticulously, where possible edge locations are, and which one or ones should be selected to deploy MEC server. MEC requires more than just deploying a server at a given edge location. This is crucial to ensure that selected edge functionality location brings both performance improvement and cost saving to the service provider. To fully harness the benefits of moving some processing and storage capabilities to the edge an evaluation of possible scenarios should be done. If the edge location is too close to the end user, MEC may become overly expensive and complex. If the edge location is too close to the centralized core, the benefits dissipate, leading to more complexity but no gain in term of performance. There are several locations where deployment of MEC functionalities could be a clever step. Inspecting the tradeoffs that these locations offer is essential for successful deployment. It is important to choose the edges location within the network painstakingly to prevent unnecessary investment. For example, Location-based applications may be housed in an aggregation point that covers the venue. An enterprise deployment may be hosted in a location that reaches all the enterprise's buildings or a part of them. For applications requiring video caching, there is more flexibility in choosing the MEC location. Since service providers may pick the location based on subscriber usage trends and see where the caching contribution could be maximized. For the abovementioned applications, stationary edge location through time may be preferable. But for a network with fluctuating loads, dynamic edge shifting according to real-time network conditions is possible and desirable in a virtualized environment. To achieve this, huge orchestration capabilities are needed in the core network. On one hand, MEC links the RAN with the UE. On the other hand, links the applications with content. It gives also the opportunity to manage traffic in real time based on factors such as application, content requirements, and network conditions to optimize available network resources use [15].

3.3 Edge Location tradeoffs

Moving processing and storage to the far edge would improve considerably the QoE and performance. However, there are many constraining parameters such as cost, complexity, space and latency on which

the choice of MEC location is based. To select the best location the choice should rest on a compromise between the criteria cited above.

Table I, shows how the MEC location impact the network key parameters as it gets closer to the UE and the opposite as it moves toward the centralized cloud:

Table 1. Impact of MEC location on network performance

Network constraining parameters	MEC location	Closer to UE	Closer to centralized cloud
Latency		Low	High
Storage Capacity		Limited	High
Processing power		Limited	High
End to end complexity		High	Low
Resources optimization		Low	High
Covered area Size		Small	Big

As the MEC location gets closer to the UE, the latency gets lower and this improves performance and QOE especially for time sensitive applications. Hardware has less storage capacity, limiting the amount of content that can be stored at the edge. Processing power becomes limited, and hence it may not be suitable to run some applications from the edge. End-to-end network complexity increases since network operators have to deploy, integrate and manage more hardware, at a higher number of locations. If applications could be smoothly run from a centralized location resources allocation at the edge may be pointless. Another key factor is the covered area. As the edge host moves away from the centralized cloud where mobility management and RAN optimization are processed, the covered zone gets smaller. Edge traffic optimization is adapted to this area. So, geographically distributed MEC servers can optimize their respective covered areas, but in separated way. But if only one MEC server covers the whole area, optimization can be coordinated across the entire area. This coordination makes optimization more efficient, but at the same time, it increases the distance to the RAN which may collapse the optimization capabilities. Access to edge applications and services is preserved as subscribers move from one area to another within the same network. If the edge host covers multiple networks, application access can be preserved across networks. As the subscriber moves to an area that has not the same MEC capabilities or doesn't support MEC at all, the subscriber may experience degradation in the quality of the access to the edge application or lose connectivity as they move out of the covered area. In many cases, this last aspect is desired especially for actors that want their services to be available only within a specific area. When the selected MEC location moves toward the centralized cloud the inverse holds giving high latency, more storage capacity, big processing power, less complexity, more efficient use of resources [15]. Table II, presents factors that play a major role in the selection of MEC location:

Table 2. Key factors defining the choice of MEC location

Factors	How MEC address the network needs
RAN resources	The MEC capacities must be adequate to the RAN capabilities to serve the covered area.
Backhaul resources	MEC may address capacity and/or latency limitations in the backhaul and prevent the backhaul from becoming a performance bottleneck.
Applications	Latency, processing and storage requirements that affect MEC location vary across applications, so the best MEC location should vary by application.
Subscriber & client expectations	The expectation for service performance and QoE may be different for enterprise employees, mall visitors.
Operator preferences	Operators may want to position themselves in the market in a specific way e.g., provide a higher-quality service for a specific enterprise client or venue owner.

4 Conclusion

In this paper, we have seen how MEC allows content, services, and applications to be accelerated by increasing responsiveness from the edge. The impact on latency, QoE and performance from moving processing and storage to the edge is easy to grasp, even though it is not trivial to quantify over a network because it depends on multiple environmental factors that are variable. We discussed also the MEC use cases that have attracted attention and their impact on network performance then we shed the light on the importance of the edge location along with the tradeoffs to be considered while deploying MEC servers across the network.

This paper is a solid basis that paves the ground to our future work that will propose the integration of MEC, DC-RAN, NFV, SDN and Big Data Analytics in an overall network architecture and discuss the benefits brought back by this combination. Furthermore, as an extension of this work, we also think that we can reuse the constraint programming concepts as proposed in [16 - 23] to bring better approaches to all these systems.

REFERENCES

[1] [1]INTERDIGITAL, “Putting Together the Pieces of the 5G Puzzle,” http://www.interdigital.com/ : White paper, October 2016, p.2.

[2] Cisco VNI Mobile, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 20162021,” http://www.cisco.com: White paper, February 2017, pp.1-28.

[3] G. A. Lewis, “Mobile computing at the edge (keynote),” in Proceedings of the 1st International Conference on Mobile Software Engineering and Systems. ACM, 2014, pp. 69–70.

[4] N. Sprecher, “MEC Activity Report 2015,” https://portal.etsi.org/.

[5] M. Patel, J. Joubert, J. Ramos, N. Sprecher, S. Abeta, A. Neal “Mobile-Edge Computing – Introductory Technical White Paper,” https://portal.etsi.org/:White paper, September 2014.

[6] D. Sabella, N. Sprecher, M. Patel, V. Young, Y. Chao Hu “Mobile Edge Computing a key technology towards 5G,” http://www.etsi.org/: Whitepaper, September 2015.

[7] E. Borcoci, “Fog Computing, Mobile Edge Computing, Cloudlets - which one?.” https://www.iaria.org/: SoftNet 2016 Conference, August 2016.

[8] INTERDIGITAL, “What the MEC? An Architecture for 5G,” http:// www.interdigital.com/ : White paper, December 2016.

[9] J. O. Fajardo, F. Liberal, I. Giannoulakis, E. Kafetzakis, V. Pii, I. Trajkovska, T. M. Bohnert, L. Goratti, R. Riggio, J. G. Lloreda et al.,“Introducing mobile edge computing capabilities through distributed 5g cloud enabled small cells,” Mobile networks and applications, vol. 21, no. 4, pp. 564–574, 2016.

[10] O. Salman, I. Elhajj, A. Kayssi, and A. Chehab, “Edge computing enabling the internet of things,” in Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on. IEEE, 2015, pp. 603–608.

[11] 5Gppp, “5G Vision,” https://5g-ppp.eu/: February 2015.

[12] M. Chen, Y. Zhang, L. Hu, T. Taleb, and Z. Sheng, “Cloud-based wireless network: Virtualized, reconfigurable, smart wireless network to enable 5g technologies,” Mobile Networks and Applications, vol. 20, no. 6, pp. 704–712, 2015.

[13] G. Klas, “ETSI MEC Use Cases and Requirements,” http://www.etsi.org/: October 2016.

[14] N. Sprecher, “Mobile Edge Computing An enabler for enhanced Car2X communication,” http://www.etsi.org/: June 2016.

[15] S. Fili, M. Paolini, “Power at the edge processing and storage move from the central core to the network edge,” : White paper, February 2017.

[16] Abourezq, M., Idrissi, A. Introduction of an outranking method in the Cloud computing research and Selection System based on the Skyline. Proceedings - International Conference on Research Challenges in Information Science. 2014.

[17] Idrissi, A. Some methods to treat capacity allocation problems. Journal of Theoretical and Applied Information Technology. 2012.

[18] Idrissi, A., Abourezq, M. Skyline in cloud computing. Journal of Theoretical and Applied Information Technology. 2014.

[19] Abourezq, M., Idrissi, A. A cloud services research and selection system. Proceedings of International Conference on Multimedia Computing and Systems. 2014.

[20] Yakine, F., Idrissi, A. Energy-aware topology control and QoS routing in ad-hoc networks. Procedia Computer Science. 2015.

[21] Idrissi, A., Rehioui, H., Laghrissi, A., Retal, S. An improvement of DENCLUE algorithm for the data clustering. 5th International Conference on Information and Communication Technology and Accessibility, ICTA 2015. 2015.

[22] Rehioui, H., Idrissi, A., Abourezq, M., Zegrari, F. DENCLUE-IM: A New Approach for Big Data Clustering. Procedia Computer Science. 2016.

[23] Idrissi, A., Yakine, F. Multicast routing with quality of service constraints in the ad hoc wireless networks. Journal of Computer Science. 2014.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Serious Games for the Development of Learning

Loubna El Azizi, Aziz Arbai
Department of mathematics. Laboratory: Algebra and Operator Theory, Faculty of Sciences
UniversityAbdelmalek Essaadi
Tetouan, Morocco
loubna.elazizi@outlook.com , arbai_aziz@yahoo.com

ABSTRACT

With the advent of interactive media, learning by playing is taking a variety of forms. Serious game, this term may seem contradictory, in reality it reflexes a very current reality, it rests on the video culture playful learners. This type of games is appropriate for investing in classrooms, the idea is to make learning more interactive and motivating. They have been developed to enable learners to build their own knowledge, skills and know-how to develop high-level skills to deal with complex situations.

In this paper, we present the concept of serious game, and our educational software who contain the attractive exercises and a serious game for mathematics.We design this softwarein Arabic to make mental calculation appreciate to novicestudents. And we believe that this type of software can be used to increasestudents'motivation to learn, and their interest in mental calculation.

Keywords-component; Serious Games; Education; Game; Learning; Video game

1 Introduction

The various works of psychologists have shown that man uses play in a natural way, in all areas of his life, as a pedagogical means to contribute to his emotional, sensory, motor, cognitive, intellectual and social development. So the introduction of the game in an educational context is not a new idea. But what is new is the relationship between learning and the numerical practices of learners.

In the 2000s, the concept of Serious Game is required to propose new applications based on the playful video culture of the learners.

In his research, Marc Prensky, explained that young people prefer images to texts, multitasking processes, networking and especially: "They prefer games to" serious "work. (Prensky, 2001), think and act differently from previous generations. Therefore, he recommends that pedagogies based onvideo games that can motivate learners and evoke serious games (Prensky, 2003).

In 2003, James Paul Gee takes up these arguments and develops them and demonstrates that video games make learning more effective and enjoyable. They encourage active, contextualized learning. They promote learners' commitment, their intuitiveness.

Over the past few years, serious games have been used in many sectors such as education, health, defense, industry, civil security and science. And they take an increasingly important place in the educational

DOI: 10.14738/tmlai.54.3216
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3216

sphere. Our goal is to design an educational software in Arabic that contains interactive exercises in mathematics and a serious game designed to make mental calculation appreciate to novice and apprentices' students, after that we will integrate this game in a pedagogical scenario and we will Determine the impact of integrating this type of game on the learning development of learners.

In this article, we present the concept of serious games, determining their definition, their history, their categories, and their characteristics, and then we determine the impacts of integrating this type of game in class and the problems posed a priori by its elaboration, after we address the problematic which led us to develop this new type of digital resource "Serious Game" for the learning of mathematics, and especially the mental calculation. In the end, we conclude by a conclusion as well as our future work.

2 Why the Serious Games ?

The emerging use of serious games poses a number of questions, the first of which relates to the definition of the expression "Serious Game", it history, it categories, and it purposes.

2.1 What is a Serious Game ?

Definition:

Several definitions of serious game have been proposed by various organizations and authors:

Definition of Zyda (2005): American researcher who participated in the development of "America's Army", specifies that the Serious Game is:

"A cerebral challenge, played with a computer according to specific rules, which uses entertainment as an added value for training and training in institutional or private environments in the fields of education, health, Civil security, as well as for communication strategy purposes. "

Definition of Julian ALVAREZ (2007): French Researcher

"Computer application, the aim of which is to combine at the same time serious aspects (Serious) such as, but not limited to, teaching, learning, communication, or information, with ludic springs coming from the Video game (Game). Such an association, which takes place through the implementation of a "pedagogical scenario", is therefore aimed at moving away from mere entertainment.

Definition of CERIMES: Center for resources and information on multimedia for higher education.

"A real training tool, communication, simulation, [the serious game is] a sort of useful declination of the video game at the service of the professionals. "Serious Games are applications developed from advanced video game technologies, using the same design and know-how approaches as the classic game (3D real-time, simulation of objects, Individuals, environments ...) but which go beyond the sole dimension of entertainment. "

From this different definition, we can conclude that the Serious game is a game that combines a serious intention with playful springs of video games. So we take a video game and we associate a utility function with it. The Serious Game must have three utility functions: Dispense a message, provide training and allow data collection.

Serious Games = Serious dimension + Videoludic dimension

2.2 History of Serious Game :

The history of the Serious Games goes back a decade. The Renaissance humanists used the oxymoron "Serio Ludere". This term refers to the idea of dealing with a serious topic with an amusing approach. So, the first traces of studies on learning through play have been in Italy among humanists.

"Called" Serio Ludere ", this approach, which could be translated as" play seriously ", would refer in particular to the notion of using humor to pass serious notions. " Julian Alvarez (2007).

As early as 1960, Business Games were introduced in universities, schools and businesses, in particular to encourage the learning of managerial techniques. The teaching technique consists of using role-playing games or modified sets of trays to train learners to react well in typical situations. It was also during this period that the term Ludo Educative (Edutainment) appeared to designate the playful activities that could be used to channel children's attention in air-conditioned centers, museums and nursery schools.

In 1970, the expression Serious Game was first used in the United States by the American researcher Clark Abt in his book entitled Serious Game. This researcher sees in the board game, the outdoor game, the role play and the game on computer, supports to spread educational, political or marketing messages.

It is also from this date, and thanks to technological advances in computer science, that the concept of Serious Game extends to video games.

Among the first Serious Games developed, one can quote OperationFrog (1983), three years later, the concept arrives in France with in particular the game “Le Sida“(1986). Other "retro serious games" are listed on the site Serious Game Classification from the research work of Alvarez and Djaouti.

But the theoretical reflection preceded the appearance of the first products, because the first true Serious Game was developed on July 4, 2002 in the United States, by and for the American army. “America's Army”. This game is a new turning point in the field of computer-based learning. It is a very realistic military simulation video game worthy of commercial games. Its free and large-scale dissemination was intended not only to reveal the reality of military operations, but also to enhance the image of the army.

After the great success of "American Army" The sector of Serious Games has evolved greatly.

2.3 The categories of Serious Game:

The After the great success of "American Army" The sector of Serious Games has evolved greatly. This diversity has led to a diversification into Serious Games families.

According to research carried out by researchers in this field, there are several classification methods.

2.3.1 A Classification of Serious Games according to the purpose:

In 2006, researchers Julian Alvarez and Olivier Rampnoux proposed to classify serious games into five main categories:

2.3.2 A Classification of Serious Games by Gameplay:

In 2007, the two researchers Ben Sawyer and Peter Smith proposed a different classification system, introducing the criterion of utility intent of the Serious Game with that of the associated market segment.

Since 2008, their classification system has been enriched with a third parameter: The Gameplay. Introduced by Damien Djaouti, Julian Alvarez and Olivier Rampnoux, he gives birth to a model called G / P / S (Gameplay, Allows, Sector):

Gameplay: This criterion notably indicates whether the Serious Game presents the possibility of winning or not. If it is not possible to win, the Gameplay is based on a video toy. This is called a Serious Play. If it is possible to win, then we are in the presence of a Serious Game based on a video game. This criterion also makes it possible to identify all the rules of the game put in the presence.

Allows: This is to list the "serious" functions associated with the videoludic base. Three major functions have been identified to date:

- Spread a message: The Serious Game aims to broadcast one or more messages. These can be of four different types: educational (ex: Edugames), informative (ex: Newsgames), persuasive (ex:Advergames) and subjective (ex: Militant games, Art games). A single game can accumulate several types of message.
- Practice: The Serious Game aims to improve the cognitive or physical abilities of the player (ex: exergames)
- Encourage the exchange of data: The Serious Game is designed to promote the exchange of data between players, or between the broadcaster of the game and the players (ex datagames).

Sector: this criterion gives two types of information:

- The scope of application of the Serious Game. For example: health, education, business, religion ...
- On the target audience of the Serious Game. For example: general public, professionals, students ... The ages of age are also to be specified.

So, we can classify the Serious Games either by the theme: (advertising, defense, politics, health, religion, ecologies) or by type, using the G / P / S model (advergame, persuasive game, newsgame, datagame).

2.4 The characteristics of Serious Games:

Usually, games share five specific characteristics.

They are goal oriented. Learners must follow specific rules while playing them. They have some sort of feedback system, and there is often an element of competition. Finally, they are voluntary, meaning that learners decide whether they like to play or not.

The Serious Game can have all the elements that characterize any other game. It can be goal-oriented with specific rules or a feedback system and a competitive comparative element. The only characteristic that is sometimes not present in the Serious Games is the voluntary character. Since these games are usually used in an educational setting or a training situation, players will often be invited to play some simulation or game.

3 Why integrate the serious games into the school environment?

As we have already pointed out, the introduction of play in the educational environment is not a recent idea. This process is taken up by teachers in nursery schools and primary to teach all subjects. This pedagogical approach aims to use the attractiveness and commitment inherent in the game to catalyze the learner's attention and engage him in his own learning (Dondlinger, 2007). Through play, the learner becomes the central actor of his formation, contrary to the passive position he occupies most of the time in traditional teaching.

He finds himself involved emotionally in the game, which facilitates the impregnation of his actions and decisions in his memory. The play can thus be advantageously used for facilitating the learning of certain skills for which traditional teaching methods are unsatisfactory (Federation of American Scientist, 2006; Mayo, 2007).

Despite the many advantages of learning by playing, this approach is today very little used in teaching. But with the arrival of new generations of learners accustomed to video games, this trend is about to change, the nature of the implicit didactic contract that prohibits learners from having fun in class as described by Chevallard (1999) is no longer appropriate for new learners.

Numerous research has shown that Games can be more effective than traditional methods in the acquisition of complex skills in higher education and in the workplace.

The integration of this type of game with the school environment allows the learner to train to perform a given task or maneuver and to study a real phenomenon reproduced in a virtual environment, learning more interactive and motivating.

Some studies also indicate that this type of game can sometimes lead to gains in confidence and concentration.

It is therefore clear that the Serious Games bring a great added value to the learning process.

4 Designing an educational software

Designing and developing an educational software and a serious game requires a pedagogical motivation. For us, it resided in the recognition of recurrent difficulties of the students in the learning of mathematics, especially in mental calculation.

There are some software and gamesadaptable to different types of these difficulties for primary students, but we decide to developa different educational software that contains interactive exercises and a serious game.

In this part, we talk about what types of games would be compatible with mental math activity. Then, we compare our software with the other.

4.1 Problem: mental calculation

Several students have difficulty in calculating the four arithmetical operations, and especially in mental calculation. And it is known that the calculation is a central activity in in mathematics and not negligible in most scientific subjects. Thus, arithmetic operations are taught through different levels of primary, colleague and secondary school.

This learning proves to be difficult, especially for novices. Students are confronted with a set of epistemological obstacles.

Several approaches are proposed in the literature to help in the resolution of these difficulties. The first is to design exercises that meet certain criteria of relevance such as being related to real world problems, being interesting, allowing creativity and innovation. A second approach uses competition to stimulate participants' skills. A third approach is to design attractive and playful environments. Finally, the latest approach uses video games to motivate students.

So, our approach, we chose to develop an educational software that contains interactive exercises to master the calculation of four arithmetic operations and a serious game in mental calculation.

4.2 Our educational software:

To help students solve the difficulties in mental calculation and to motivate them, we design an educational software containing interactive exercises in algebra, and a serious game in mental calculation. It targets novices'students in the 6th year of primary school.

There are several types of video games that seemed to be representative of the gaming sphere: combat games, action-adventure games, action-RPG games, shooting games, sports games, games Racing games, electronic matching games, arcade video games, multiplayer Online Battle Arena games, strategy games, role playing games, platform games, puzzle games, adventure. After the analysis of each type of video games, we chose the action-RPG video game.

The principle of this game is that the player usually embodies a single character and the game happens in real time. The player evolves freely in a world and has certain goals to accomplish to advance in the adventure. In addition, most A-RPGs offer a system of advanced character evolution (equipment, characteristics, level of experience, etc.). All movements are on the same plane and the enemies are directly visible. Fighting begins when the enemy is within reach and takes place in the same environment.

They usually interpret a character moving in a 3D universe to carry out different types of actions: discovering secret zones, collecting objectsetc., or fighting enemies, knocking out invaders, the scenarios can be various and varied.

As opposed to other existing software or games, our software is developed in Arabic, and can be used to acquire new concepts, in the consolidation phase or in the evaluation phase. So according to the pedagogical choice of the teachers. And from the interactive exercises, the studentwill learnthe methods for calculating addition, subtraction,

multiplication and division of integers and decimals numbers, and fractions. And He will develop their ability to mental calculating using our serious game.

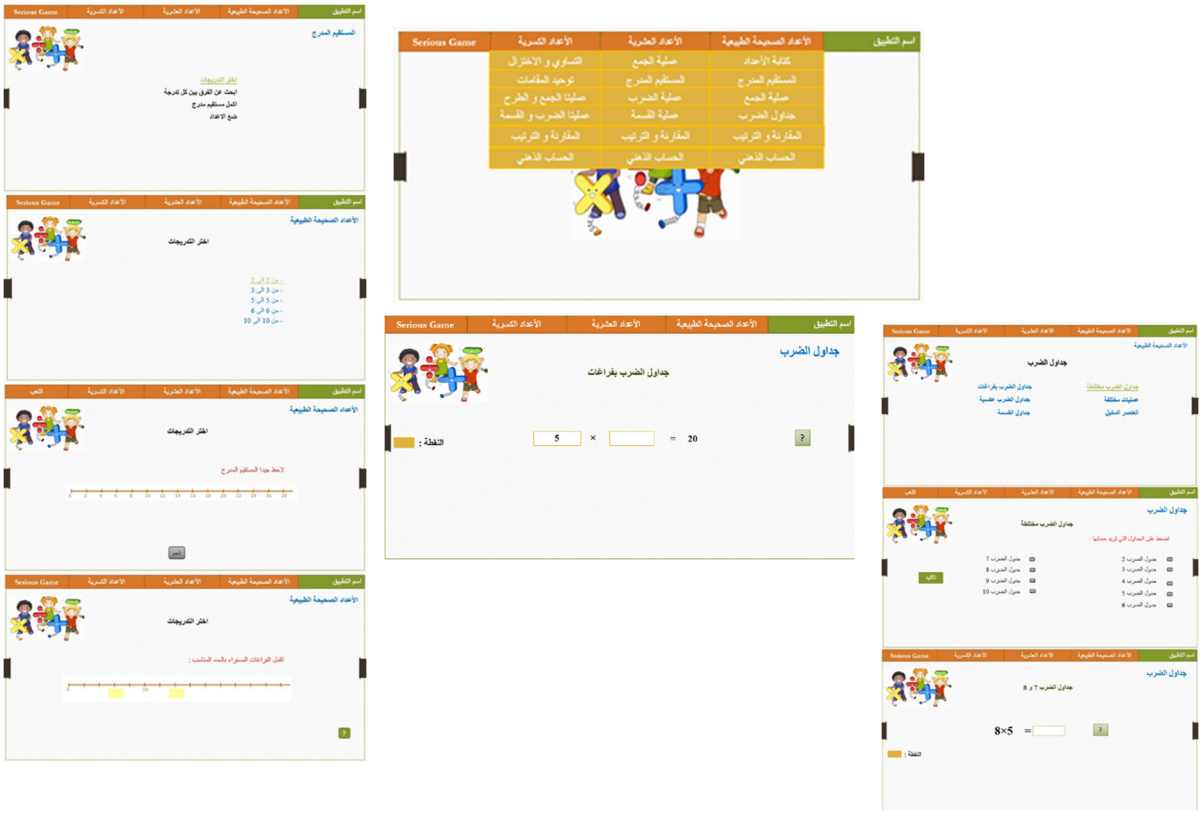
Therefore, the interactives activities and the serious game allow the students in difficulty to be more motivated and therefore to become more involved in the activity.

We can summarize the goals to be achieved to this software in two:

The utilitarian aims are: To help the student in his school life, knowledge of the numbers and the processes, mental gymnastics.

The educational goals are: Developing auditory memory, attention, self-control, deploying a synthetic activity, keep the whole problem in mind when dealing with a part of it.

4.3 Software screenshot:



5 Conclusion

The concept of serious games intended to support learning through the practice of the game has been a great success in various fields, for which the pedagogies based on the use of serious game have expanded widely. They are in fact capable of soliciting the motivation of the learners and of enabling them to develop knowledge in the context of complex learning situations. For this we decided to develop educational software contains interactive exercises and serious game, in order to use the numerical practices of learners to develop their skills in mental calculation.

Several researches have been launched in the experimentation of these new learning tools. For this we will integrate our educational software into an educational scenario, and we will do an experimentation in primary school, and we will determine the impact of this educational software on the learning development of learners.

REFERENCES

[1] [1]ALVAREZ, Julian, « Du Jeu vidéo au Serious Game : approches culturelle, pragmatique et formelle », Thèse de doctorat en science de l'information et de la communication, Université de Toulouse II le 17.12.2007.

[2] Damien DJAOUTI , «Serious Game Design Considérations théoriques et techniques sur la création de jeux vidéo à vocation utilitaire» Thèse de doctorat en Informatique, Université Toulouse III Paul Sabatier le 8.11.2011.

[3] ALVAREZ Julien, DJAOUTI Damien, Introduction au seriousgame, Ed. Questions théoriques, 2010.

[4] SANCHEZ, Eric., et al., "Jeux sérieux et pédagogie universitaire : de la conception à l'évaluation des apprentissages". Revue internationale des technologies en pédagogie universitaire,(2011). (en ligne : http://www.ritpu.org/IMG/pdf/RITPU_v08_n01-02_48.pdf).

[5] Zyda, M., « From Visual Simulation to Virtual Reality to Games»,IEEE Computer, vol. 38, n° 9, 2005, p. 25-32.

[6] Viau, R., La motivation en contexte scolaire, Bruxelles : De Boeck,1997.

[7] Serious Game classification, site créé en 2006 par LudoScience (Julian Alvarez, Damien Djaouti en association avec les chercheurs des laboratoires IRIT et LARA des universités de Toulouse II et III). Disponible sur : serious.gameclassification.com.

[8] Sébastien George et Éric Sanchez, Actes de l'Atelier « Jeux Sérieux : conception et usages ». 4ème Conférence francophone sur les Environnements Informatiques pour l'Apprentissage Humain. 23 juin 2009.

[9] Cobb, P., Confrey, J., diSessa, A., Lehrer, R., Schauble, L. (2003). Design experiments in educational research. Educational Researcher, 32, 1: 9 – 13.

[10] Malone, T. W., Lepper, M. R., « Making learning fun: Ataxonomy of intrinsic motiavations for learning », Aptitude, learning andinstruction: III. Conative and affective process analyses, 1987, p. 223-253.

[11] Jeux sérieux / Seriousgames, site édité par le CERIMES (Centre de Ressources et d'Information sur les Multimédias pour l'Enseignement Supérieur). Disponible sur : http://www.jeux-serieux.fr/

[12] « Game Based Learning, Apprendre avec les jeux vidéo » Institut suisse des médias pour la formation et la culture, décembre 2013 http://guides.educa.ch/sites/default/files/educa.guides_gamebasedlearning_fr_0.pdf.

[13] L'historique des rapports entre pédagogie et jeu : http://histgeo.discip.ac-caen.fr/ludus/histjeduc.htm.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Novel Compact CPW LowPass Filter Integrating Periodic Triangle DGS Cells

[1]Elmahjouby Sghir, [1]Ahmed Errkik, [1]Jamal Zbitou, [1]A.Tajmouati, [1]Larbi El Abdellaoui, [2]Mohamed Latrach

[2]LMEET Laboratory FST of Settat Hassan 1st University, Settat, Morocco
[2]Microwave group ESEO, Angers, France
elmahjouby.sghir@gmail.com

ABSTRACT

In this paper, we introduce a new periodic structure for CPW of a low pass filter based on the DGS technique with triangle slot cell forme. The proposed structure is a minuature low pass filter that exhibits low insertion loss in the passband, simple fabrication, and easy to integrate with microwave planar circuits.

A Lowpass filter (LPF) has been designed with flat response at a selected frequency by cascading three DGS cells.

Keywords—cpw lowpass filter; triangle DGS; periodic cells; wide passband;

1 Introduction

Periodic structures of various types in coplanar wave guide (CPW) technologies has a potentially high volume demands in microwave and millimetre-wave domains to their filtering properties or inhibition of signal propagation in certain frequency bands. Also, Defected Ground Structures (DGS) have been developed to improve response and characteristics of microwave components. So when both of them combined in the same structure will give excellent performance in terms of response as bandwidth, quality factor and the parasite rejection, also the level of integration as the size, the development and integration of other components. The microwave circuits such as filters design[1], antennas [2], phase shifters [3], and other are of much application domain of CPW with DGS structure.

All this attractive of CPW-DGS structures because of useful for fabrication, characterisation and manufacture of a small, low-cost and highly integrated transceiver. However, promising application of CPW in microwave integrated circuits (MICs), monolithic microwave integrated circuits (MMICs), and microelectromechanical systems (MEMS) devices[4, 5].

In transmission and reception channels of modern microwave systems, such a wireless and mobile communications, the harmonics and spurious responses are caused by the caractiristic response of some no-linear devices like oscillator. So, it is necessaire to suppress harmonics and spurious signals in the bandpass signal. For this purpose, a compact CPW Low-pass filter (CPW-LPF) using DGS periodics structure is a good condidate to improve passband performances.

DOI: 10.14738/tmlai.54.3218
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3218

In this paper, a novel compact low pass filter meeting these requirements (CPW , DGS , Periodic cells) with high performances : Harmonic-suppressed, sharp attenuation and wide stop-band.

2 Initilal DGS LPF Design

2.1 CPW line based on one Triangle DGS

The geometry of the proposed Low Pass filter using Triangle DGS filter is depicted in Figure 1. The design started with one conventional cell, the DGS configuration in this structure is an Isosceles triangle slot etched in the ground

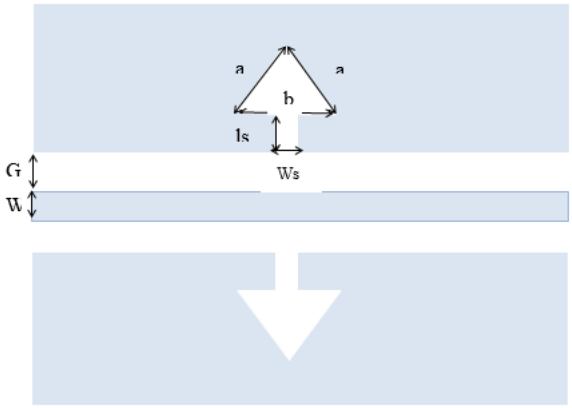


Figure. 1. CPW line with Triangle DGS structure

The coplanar waveguide line is designed to have a charactiristic impedance Z0=50 Ω, and built on standard 25-mil Duroid substrate with dielectric constant ε r = 10.2,[6]. The central signal line width of W=1.2 mm, and the gap width is G= 0.42 mm. Also the parameters of the triangle DGS are a=1.1181 mm, b=1.0004 mm, and ls=ws=0.42 mm. The overall dimensions of the simulated device measures 20×14.04 mm2. The parametric studies and simulation of the filter's electrical performances were performed using Momentum (an EM solver integred in ADS Agilent).

2.2 Response of CPW DGS Triangle

The Tringle DGS element etched in the ground plane has the effective capacitance (C) and inductance (L). The simulation insertion and return losses of the proposed topoligie filter is depicted in Figure 2.

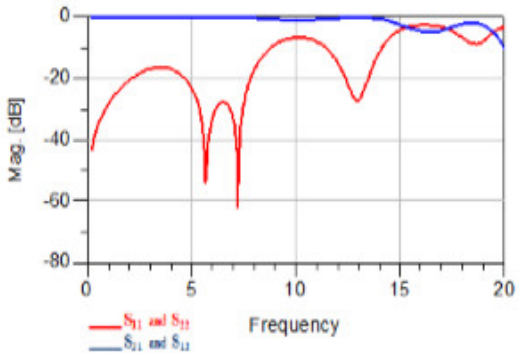


Figure. 2. Simulated S-parameters versus frequency for the CPW DGS Triangle structure.

From the S-parameter responses, it is clear that the filter provides low pass band behavior, with large bandpass from 0.1 to 15 GHz, very low and flat insertion loss of –0.5 dB and the minimum return loss of –10 dB in passband, but the suppression is inadequate because limit on the separation between the insertion loss and return loss in the stopband.

3 Proposed Device of LPF with Periodic DGS Cells

3.1 CPW line based on Three Triangle DGS Cells

In the research, augmentation of DGS elements number were used as a good solution to design a compact microwave filter, [7, 8, 9]. The performance of LPF with periodical DGSs was used here to have a much wider stopband compared with one DGS. Figure 3 depicts the schematic of the proposed structure with periodic cells

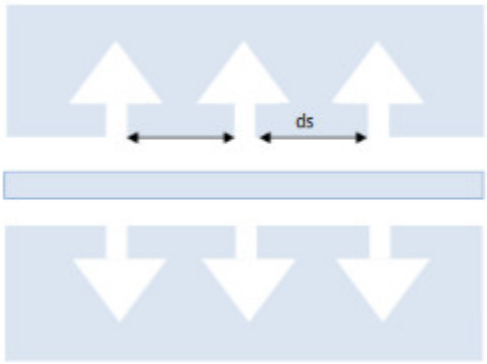


Figure. 3. The proposed device of periodics DGS Cells for LowPass filter

The parameters of the initial triangle DGS low pass filter are kept unchanged while two other symmetrical triangle DGS are introduced and separed with ds=2.9834mm distance.

However, the insertion-loss and return-loss performance were inadequate. As a result, a LPF with three cells periodical triangle DGSs was designed in an attempt to improve the performance of S-parameters as shown in Fig.4.

Figure 4 shows the simulated S-parameters of the proposed device of low pass filter with three periodic Cells

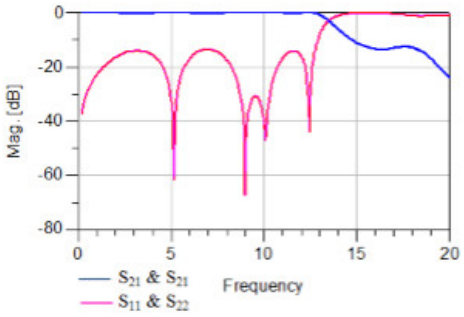


Figure. 4. Simulated S-parameter of the proposed periodic DGS cells device

As a result shown in Fig. 4, the cascaded LPF with the three periodic cells present a sharp roll-off with improved stop band and supression performance. The return loss in pass-band is greater than 15 dB, with

no more than 0.3 dB ripple level in the pass-band, while the stop band rejection level is better than 10 dB with wide stopband of 13 to more than 20 GHz.

3.2 Effect of additional periodic cells on filter responce.

The simulation results of three Triangle DGS cells are shown in Figure 4. It is clearly shown in the figure that three-cell in the structure's ground is enough to suppress the spurious response in pass and stop bands and give out a wider stopband. Figures 5 and 6 demonstrates the effect of additional periodic cells on insertion and return loss.

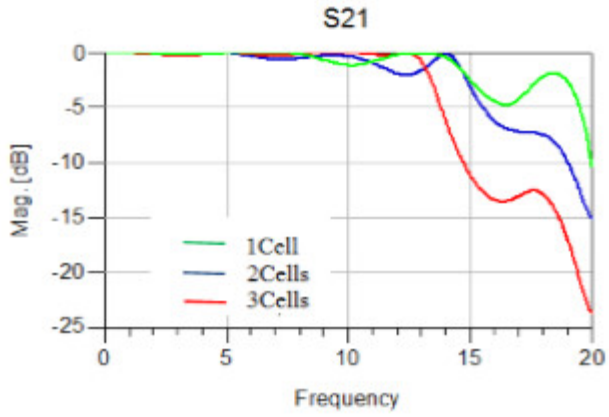


Figure. 5. Effect of additional periodic cells on filter insertion loss

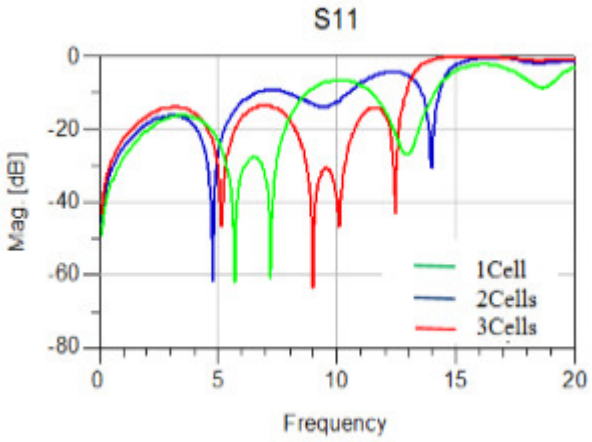


Figure. 6. Effect of additional periodic cells on filter return loss.

Figures 5 and 6 demonstrated that the augmentation of periodic tiangle DGS cells take a great effect on the filter response. There is a good acheiving of attenuation with no ripple level of insertion loss in the pass-band with just 3 cells. Also, the filter present spurious-free response and deep attenuation levels in the return loss -60 dB in passband and rejection level better than -20 dB in the stopband, when the number of DGS cells augmented.

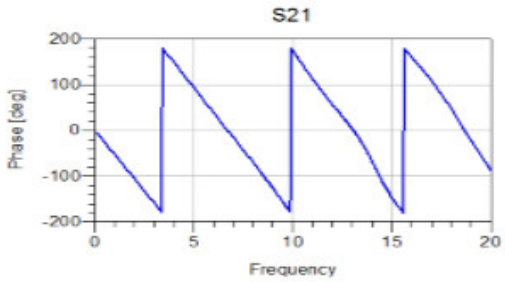


Figure. 7. Simulation Phase response of Lowpass Filter Designed.

Figure 7 shows the phase simulation of insertion loss (S21) for UWB bandpass between 0.1 GHz and 20 GHz of designed lowpass filter that throughout the pass and stop bands is acceptably linear for UWB applications.

4 Conclusion

In this paper, a simple compact lowpass filter using triangle defected ground structure is proposed, developed and presented. With 3-Cells DGS series on the same substrate offers the advantages of miniature low-pass filter than a standard CPW transmission line, low insertion loss with no harmonics in the passband, high attenuation levels in the stopband and very easy fabrication, while the cutoff frequency can be adjusted by controlling the dimensions of the triangle DGS cell to obtained others response frequency for more wireless applications.

ACKNOWLEDGMENT

We thank Mr. Mhamed Latrach from ESEO Angers in France, for allowing us to use and to perform simulations by using electromagnetic solvers and measurement instruments

REFERENCES

[1] Hettak, K. and G. Delisle, “A new miniature uniplanar lowpass Filter using series resonators, ” IEEE MTT-S, Vol. 3, No. 3, 1193-1196, Jun. 7-12, 1998.

[2] Chang, I., and Lee, B. : “Design of defected ground structures for harmonic control of active microstrip antenna,”. 2002 IEEE AP-S Int. Symp. Dig., 2002, pp. 852–855.

[3] Elamaran, B, Chio, I.M, Chen, L.Y, and Chiao, J.C, “A beamsteerer using reconfigurable PBG ground plane,” 2000 IEEE MTT-S Int. Microwave Symp. Dig., 2000, pp. 835–838.

[4] S. Seki and H. Hasegawa, “Cross-tie slow-wave coplanar waveguide on semi-insulating GaAs substrate,” Electron. Lett., vol. 17, no. 25, pp. 940-941, Dec. 1981.

[5] K. Wu and R. Vahldieck, “Hybrid-mode analysis of homogeneously and inhomogeneously doped low-loss slow-wave coplanar transmission lines,” IEEE Trans. Microwave Theory Tech., vol. 39, no. 8, pp. 1348-1360, Aug. 1991.

[6] Jong-Sik Lim, Chul-Soo Kim, Young-Taek Lee, Dal Ahn, Sangwook Nam, "A Spiral-Shaped Defected Ground Structure for Coplanar Waveguide," IEEE Microwave and wireless conponents letters, vol. 12, no. 9, September 2002.

[7] J. Y. Wu, Y. H. Tseng and W. H. Tu, "Design of compact lowpass filter with ultra-wide stopband using thin slots," Progress In Electromagnetics Research C, Vol. 31, pp. 137-151, 2012.

[8] J. W. No and H. Y. Hwang, "A design of cascaded CPW low-pass filter with broad stopband," IEEE Microwave and Wireless Components Letters, Vol. 17, pp. 427-429, 2007.

[9] C. Han-Jan, H. Tsung-Hui, and C. Chin-Sheng et al., "A novel cross-shape DGS applied to design ultra-wide stopband low-pass filters," IEEE Microwave and Wireless Components Letters, Vol. 16, No. 5, 2006, pp. 252-254.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Robust Multiple Watermarking Technique for Medical Applications using DWT, DCT and SVD

Imane Assini, Abdelmajid Badri, Khadija Safi, Aicha Sahel, Abdennaceur Baghdad
EEA&TI laboratory Faculty of Sciences and Techniques (FSTM), Hassan II University Casablanca, BP 146, Mohammedia, 20650 Morocco
assini.media@gmail.com

ABSTRACT

The security of medical data plays an important role in preserving their integrity and confidentiality. This paper presents a robust multiple watermarking algorithm based on a combination of discrete wavelet transform (DWT), discrete cosine transform (DCT) and singular value decomposition (SVD).

In this proposed method, we have inserted three watermark images (a patient image, a doctor image and a fingerprint of the patient) into a single cover medical image. In the insertion phase, the original image is decomposed into the third level of DWT, and then the DCT and SVD transforms are applied on the selected sub bands (HH1, HH2, and HH3). The first watermark goes through the same steps that the original image. Moreover, the second and the third watermarks are decomposed until the second level and the first level by DWT respectively and subsequently transformed by SVD. The singular values of the three watermarks are inserted into the singular values of the original image. The experimental results show that the proposed method gives better performance in terms of capacity, quality and robustness against various attacks compared to other reported techniques.

Keywords-multiple watermarking, medical image, DWT, DCT, DWT.

1 Introduction

Telemedicine plays an important and vital role in the aid to the diagnosis and decision making; it is a well-known application that facilitates the transfer of the enormous amount of medical data[1].

Our approach is situated within this frame and it consists on conceiving a control and monitoring system in order to protect medical data spread between hospital centers. The idea was then to think to the digital watermarking. Digital watermarking is a technique of embedding information into a medical image (Magnetic Resonance (MRI), Computed Tomography (CT), Mammography (MG)…) for various purposes such as the fight against fraud and the protection of copyright. Recently, multiple image watermarking appeared as a solution to enhance the confidentiality and security of the inserted watermarks.

In the literature, there are a number of contributions using the multiple image watermarking such as:

In [2], the method proposed by the authors is based on multiple watermarking for healthcare application which combines DWT, DCT, and SVD. In the embedding process, two watermarks are inserted into the cover medical image after their decomposition up to the third level by DWT and their transformation by

DOI: 10.14738/tmlai.54.3219
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3219

DCT and SVD respectively. The experimental results show that this contribution gives an excellent performance in terms of robustness and imperceptibility. In [3], the authors present a secure multiple watermarking technique based on the combination of DWT, DCT, and SVD for medical images. This proposed method uses three watermarks, one medical image and two text watermarks. This watermarking scheme gives a good efficiency for security, capacity, imperceptibility and robustness.

In [4], the authors proposed a hybrid technique for medical images based on Non sub sampled contourlet transform (NSCT) which combines discrete cosine transform (DCT), Multiresolution Singular value decomposition(MSVD) and Arnold transform. Three watermarks (two images and text watermark) are embedded into a single cover medical image. Experimental results approve that this proposed method produces high robustness against various attacks.

The steps of our paper are as follows: Step 1 presents the techniques used in our approach, in the step 2 the hybrid algorithm DWT-DCT-SVD is proposed, in the step 3 the experimental results are presented and in the final step a conclusion of the work is done.

2 Used techniques

The proposed multiple watermarking is based on DWT, DCT, and SVD. The choice of these three techniques it's focused on their properties, DWT allows a good location in the time and the frequency domain and it's better to identify appropriate data to human perception (HVS)[3], DCT has a good perceptual invisibility, an acceptable robustness, and a reasonable complexity[2, 5]. And finally, the important properties of the SVD is that even if one applies a small variation in the singular values it does not affect the quality of the image [6] as well as the singular values of an image have a great stability.

2.1 Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) [7] is a multiresolution description which consists of decomposing image into four frequency representations. These four representations are called sub-bands LL (approximation), LH (vertical), HL (horizontal) and HH (diagonal).

2.2 Discrete Cosine Transform (DCT)

The discrete cosine transform (DCT) [8] is a function that transforms the representation of a data from the spatial domain to the frequency domain. The energy of the image is concentrated in the sub band of the low frequencies which contains the most important visual part of the image. Most watermarking techniques based on DCT embed the watermark in the middle frequency sub-bands because attacks can delete the high-frequency components of the image.

2.3 Singular Value Decomposition (SVD)

The singular value decomposition (SVD) [6] of an image is represented by a matrix which can be factored into a product of two orthogonal matrices and a diagonal matrix.

The decomposition of an image into SVD can be written as (1):

$$I = U*S*VT \quad (1)$$

Where S is a matrix whose diagonal terms are positive and orderly decreasing, and all other terms being zero. The singular values represent the energy of the image.

3 Performance measures

The Imperceptibility of the proposed watermarking system is evaluated by the calculating the Peak Signal to Noise Ratio(PSNR) (3) which is based on the mean square error (MSE) defined by (2)[9]:

$$\boldsymbol{MSE} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[\boldsymbol{X}(\boldsymbol{i},\boldsymbol{j}) - \boldsymbol{X}'(\boldsymbol{i},\boldsymbol{j})]^2 \quad (2)$$

Where X and X' are the cover medical image and the watermarked medical image respectively of size m x n.

$$\boldsymbol{PSNR} = \boldsymbol{10\,log}\left[\frac{\boldsymbol{255}^2}{\boldsymbol{MSE}}\right] \quad (3)$$

Furthermore, the robustness of the watermarking scheme is measured by the correlation between the reference watermark and the extracted watermark by (4)[9]:

$$\boldsymbol{NC} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(\boldsymbol{I}(\boldsymbol{i},\boldsymbol{j})*\boldsymbol{I}'(\boldsymbol{i},\boldsymbol{j}))}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\boldsymbol{I}(\boldsymbol{i},\boldsymbol{j})^2*\sum_{i=1}^{M}\sum_{j=1}^{N}\boldsymbol{I}'(\boldsymbol{i},\boldsymbol{j})^2}} \quad (4)$$

Where I(i,j) is a pixel of the original watermark of size M×N and I' (i, j) is a pixel of the extracted watermark of size M×N.

4 Proposed method

The proposed method DWT-DCT-SVD has two steps, embedding and extracting process illustrated in figures 1 and 2.

4.1 Embedding process

1. Decomposing the cover medical image into the third level DWT transform using Haar wavelet and then select HH1, HH2, and HH3.
2. Apply step1 to the first watermark image 'Patient image' and select $HH_{3patient}$.
3. Apply the second and the first level of DWT to the second and the third watermark images 'Doctor image' and 'Fingerprint' respectively and then select $HH_{2doctor}$ and $HH_{1fingerprint}$.
4. Apply DCT on the selected sub bands HH1, HH2, HH3 and HH3patient of the original medical image and watermark image1 'Patient image' respectively and then apply SVD on DCT of (HH1, HH2, HH3 and HH3patient) and on HH2doctor and HH1fingerprint of the second and the third watermark images to get :

$$\boldsymbol{A_i} = \boldsymbol{U_i S_i V_i^T} \quad \text{i=HH}_1\text{, HH}_2\text{ and HH}_3 \quad (5)$$

$$\boldsymbol{w_i} = \boldsymbol{U_j S_j V_j^T} \quad \text{j= Patient image, Doctor image and Fingerprint} \quad (6)$$

5. Modify the singular values of HH3 of the cover medical image with the singular values of HH3patient of watermark image 1 'Patient image', the singular values of HH2 of the cover medical image with the sigular values of $HH_{2doctor}$ of watermark image 2 'Doctor image' and finaly the singular values of HH_1 of the cover medical image with the singular values of $HH_{1fingerprint}$ by (7):

$$\boldsymbol{A_{wati}} = \boldsymbol{S_i} + \boldsymbol{\alpha} * \boldsymbol{S_j} \quad (7)$$

$\boldsymbol{\alpha}$:Scaling factor

6. Apply inverse DCT and then inverse DWT to get the watermarked medical image.

4.2 Extracting process

1. Decomposing the cover medical image into the third level DWT transform using Haar wavelet and then select HH_1, HH_2, and HH_3.
2. Apply step1 to the first watermark image 'Patient image' and select $HH_{3patient}$.
3. Apply the second and the first level of DWT to the second and the third watermark images 'Doctor image' and 'Fingerprint' respectively and then select $HH_{2doctor}$ and $HH_{1fingerprint}$

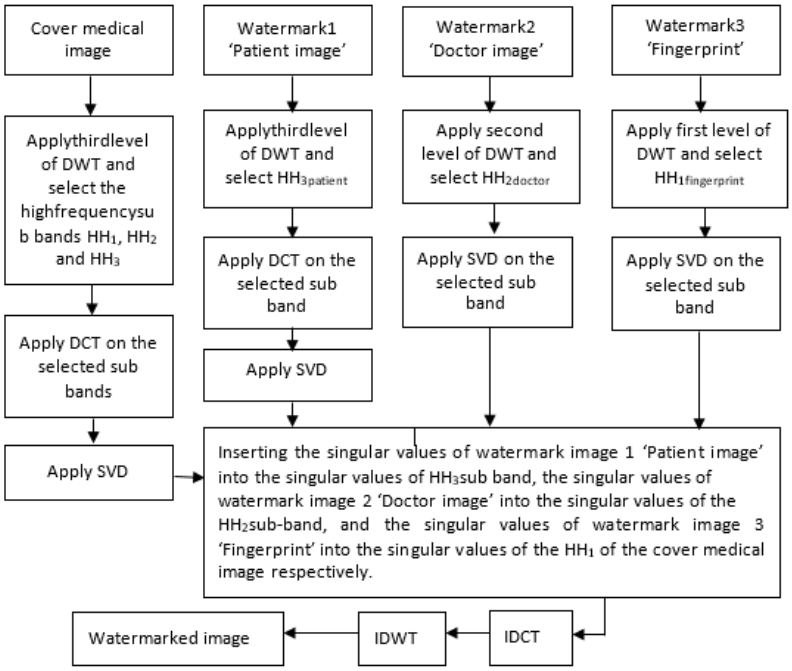


Figure 1. Embedding process using DWT-DCT-SVD

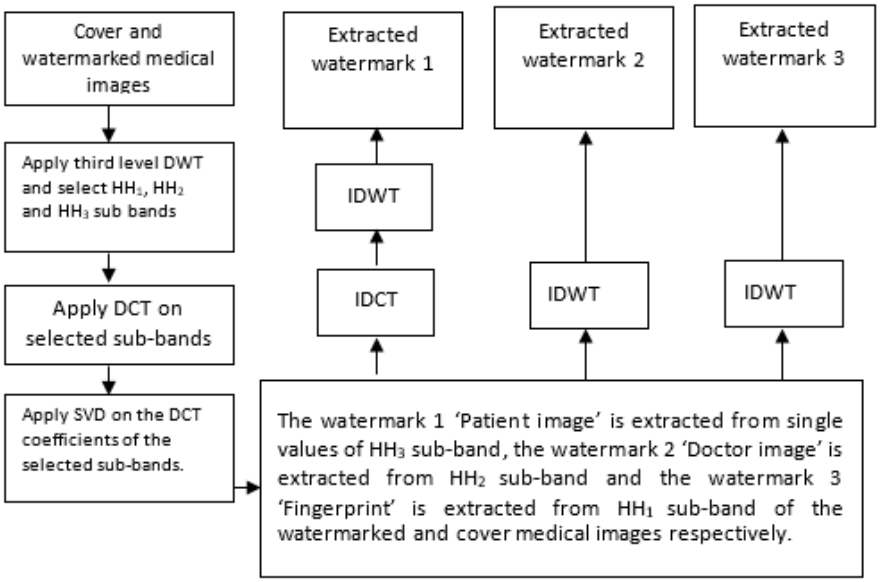


Figure 2. Extracting process using DWT-DCT-SVD

4. Apply DCT on the selected sub bands HH_1, HH_2, HH_3 and $HH_{3patient}$of the original medical image and watermark image1 'Patient image' respectively and then apply SVD on DCT of (HH_1, HH_2, HH_3 and $HH_{3patient}$) and on $HH_{2doctor}$ and $HH_{1fingerprint}$ of the second and the third watermark images to get :

$$\boldsymbol{A_i} = \boldsymbol{U_i S_i V_i^T} \text{ , i=HH}_1\text{, HH}_2 \text{ and HH}_3 \quad (8)$$

$$\boldsymbol{w_i} = \boldsymbol{U_j S_j V_j^T} \text{ j= Patient image, Doctor image and Fingerprint} \quad (9)$$

5. Apply step 1 and 4 to watermarked medical image to obtain:

$$\boldsymbol{A_{wati}} = \boldsymbol{U_{wati} S_{wati} V_{wati}^T} \text{i=HH}_1\text{, HH}_2 \text{ and HH}_3 \quad (10)$$

6. The singular values of 'Patient image', 'Doctor image' and 'Fingerprint' are obtained by the singular values of HH_3HH_2, and HH_1 of the watermarked medical image and the cover medical image respectively by (11):

$$\boldsymbol{S'_j} = (\boldsymbol{S_{wati}} - \boldsymbol{S_i})/\boldsymbol{\alpha} \quad (21)$$

7. The extracted watermark images 1 'Patient image' is obtained by applying ISVD using equation (11) and then applies inverse DCT and inverse DWT.
8. The extracted watermark image 2 'Doctor image' and the extracted watermark image 3 'Fingerprint' are obtained by applying ISVD using equation (11) and then apply inverse DWT.

5 Experimental results

The proposed approach DWT-DCT-SVD was simulated using Matlab R2013b.

a number of tests are applied on three cover medical images 'Brain', 'CT scan', and 'Knee' of size '512 x 512' and 'Patient image', 'Doctor image' and 'Fingerprint' of size 512x512 taken as watermark images shown in Figure 3.

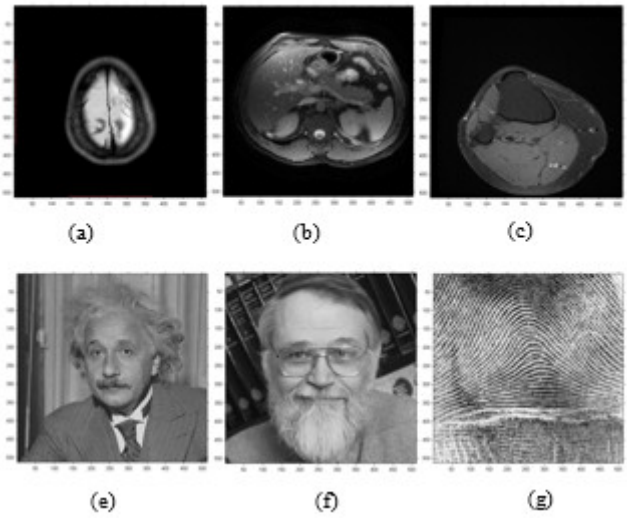


Figure 3. (a, b, c) Cover medical images: a) Brain. b) CT scan. c) Knee, (e, f, g) watermark images : e). Patient image. f) Doctor image. G) Fingerprint

To evaluate the imperceptibility of our algorithm we have calculated the PSNR between the cover medical images and the watermarked images. Figure 4 approves that the proposed method gives an excellent quality of watermarked image.

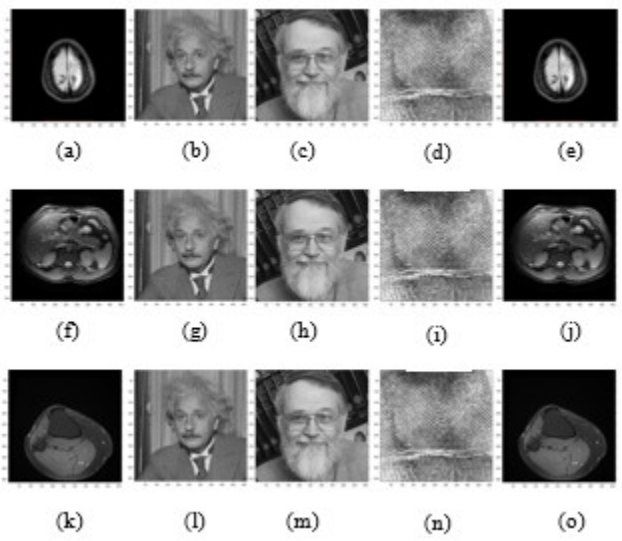


Figure 4. (a, f, k) cover medical images. (b, g, l) watermark 1 'Patient image'. (c, h, m) watermark 2 'Doctor image'. (d, i, n) watermark 3 'Fingerprint'. (e, j, o) watermarked images

The Tables 1, 2 and 3 show the quality of perceptibility of the proposed method without any noise attacks. When the cover medical image is 'Brain', The maximum of PSNR value obtained against the three watermark images 'Patient image', 'Doctor image ' and 'Fingerprint' is 42.70 at scaling factor 0.1 and minimum PSNR value is 36.84 at scaling factor 2. In the case where the cover image is 'CT scan' the maximum PSNR value is 37.30 at scaling factor 0.1 and the minimum PSNR value is 33.02 at scaling factor 2 and with the cover medical image 'Knee' the maximum PSNR is 40.11 at scaling 0.1 and minimum PSNR value is 33.53 at scaling factor 2. On the other hand, the NC values are 1 in all scaling factor shown in Tables 1, 2 and 3.

Table 1. PSNR and NC Performance of the Proposed Method for medical image Brain

Gain factor	Brain			
	PSNR	NC (Patient image)	NC (Doctor image)	NC (Fingerprint)
0.1	42.70	1	1	1
0.2	42.19	1	1	1
2	36.84	1	1	1

Table 2. PSNR and NC Performance of the Proposed Method for medical image CT scan

Gain factor	CT scan			
	PSNR	NC (Patient image)	NC (Doctor image)	NC (Fingerprint)
0.1	37.30	1	1	1
0.2	36.96	1	1	1
2	33.02	1	1	1

The PSNR values obtained by the proposed algorithm give better results than other techniques [3, 2, and 6]. At scaling factor 0.1 the maximum PSNR value is 42.70 compared to 32.09 obtained by [3], 33.08 obtained by [2] and 32.19 obtained by [6]. When scaling factor is 0.2, the maximum PSNR value is 42.19 compared to 27.29 obtained by [3] and finally at scaling factor 2, the maximum PSNR value is 36.84 compared to 27.78 obtained by [2].

Table 3. PSNR and NC Performance of the Proposed Method for medical image Knee

Gain factor	Knee			
	PSNR	NC (Patient image)	NC (Doctor image)	NC (Fingerprint)
0.1	40.11	1	1	1
0.2	39.44	1	1	1
2	33.53	1	1	1

To evaluate the robustness of our approach technique, various attacks were applying such as Gaussian noise, Speckle noise, Salt & pepper and Median filtering. Table 4 shows the performance of NC values at scaling factor 0.2.

The maximum NC value obtained against Gaussian noise, Speckle noise, Salt & pepper and Median filtering is 1 for watermark image ‘Doctor image’. However, the minimum NC value is 0.9571 against Speckle noise for watermark image ‘Fingerprint’.

Table 4. Performance of NC values at gain factor = 0.2

Attacks	Brain			CT Scan			Knee		
	NC Patient	NC Doctor	NC Fingerprint	NC Patient	NC Doctor	NC Fingerprint	NC Patient	NC Doctor	NC Fingerprint
Gaussian noise (M = 0,V = 0.0 0003)	0.995 6	**1**	0.9997	0.9988	**1**	0.9999	0.999 7	**1**	0.9998
Gaussian noise (M = 0,V =0.001)	0.995 9	**1**	0.9840	0.9985	**1**	0.9748	0.999 4	**1**	0.9728
Gaussian noise (M = 0.01,V =0.0005)	0.996 2	**1**	0.9906	0.9991	**1**	0.9880	0.999 6	**1**	0.9863
Salt & pepper (density = 0.001)	0.997 2	**1**	0.9874	0.9991	**1**	0.9904	0.999 6	**1**	0.9870
Salt & pepper (density = 0.002)	0.997 1	**1**	0.9662	0.9991	**1**	0.9735	0.999 0	**1**	0.9687
Speckle noise (density = 0.0002)	0.996 1	**1**	0.9996	0.9989	**1**	1	0.999 8	**1**	1
Speckle noise (density = 0.005)	0.997 2	**1**	0.9957	0.9993	**1**	0.9934	0.999 9	**1**	0.9968
Speckle noise (density = 0.02)	0.998 2	**1**	0.9691	0.9996	**1**	0.9571	0.999 8	**1**	0.9769
Median filtering	0.987 7	**1**	0.9979	0.9826	**1**	0.9990	0.992 6	**1**	0.9985

In Table 5, we have compared the robustness of our technique against various attacks compared to existing techniques [2], [3] and [10].

The maximum NC value obtained at scaling factor= 0.2 is 1 for watermark image 'Doctor image' against Gaussian noise compared to 0.9908obtained by [2], 0.9466 obtained by [3] and 0.944001 obtained by [10] respectively, 1 for watermark image 'Doctor image' against Salt & pepper compared to 0.9953obtained by [2], 0.9658obtained by [3] and 0.949754obtained by [10] respectively, 1 for watermark images 'Doctor image' and 'Fingerprint' against Speckle noise (density = 0.00001)compared to 0.949742obtained by [10], 1 for watermark image 'Doctor image' against Median filtering compared to 0.9861 obtained by [2], 0.0123 obtained by [3] and 0.9985 obtained by [10] respectively, 1 for watermark image 'Doctor image' against Salt & Peppers (d = 0.002) + Speckle (V=0.005) compared to 0.8328 obtained by [3], and finally 1 for watermark image 'Doctor image' against Gaussian (M=0.01,V=0.002)+ Speckle (V=0.005)compared to 0.8276 obtained by [3].

Table 5. Comparison of the proposed method with existing approaches at scaling factor 0.2

Attacks	Proposed method			[2]		[3]	[10]
	NC Patient	NC Doctor	NC Fingerprint	NC Symptoms	NC Record		
Gaussian noise (M=0, V=0.00003)	0.9997	**1**	0.9998	Not shown	Not shown	Not shown	0.944001
Gaussian noise (M=0, V=0.001)	0.9994	**1**	0.9728	0.9515	0.9908	0.9466	Not shown
Salt & pepper (density = 0.001)	0.9996	**1**	0.9870	0.9636	0.9953	0.9658	0.949754
Speckle noise (density = 0.00001)	0.9998	**1**	1	Not shown	Not shown	Not shown	0.949742
Speckle noise (density = 0.02)	0.9998	**1**	0.9769	Not shown	Not shown	0.8673	Not shown
Median Filtering	0.9926	**1**	0.9990	0.8657	0.9861	0.0123	0.9985
Salt & Peppers (d = 0.002) + Speckle (V=0.005)	0.9989	**1**	0.9615	Not shown	Not shown	0.8328	Not shown
Gaussian (M=0.01,V=0.002)+ Speckle (V=0.005)	0.9991	**1**	0.9284	Not shown	Not shown	0.8276	Not shown

6 Conclusion

This article presents an innovative multiple watermarking method based on the combination of three techniques DWT, DCT, and SVD. Three watermark images 'Patient image', 'Doctor image' and 'Fingerprint' are embedded into a cover medical image of modalities: computer tomography (CT) and magnetic resonance imaging (MRI). The insertion is made in the high-frequency sub bands HH1, HH2 and HH3 of the cover medical image. However, the insertion in these areas makes it possible to enhance the robustness of the system of watermarking without damaging the visual quality of the watermarked image. The results obtained approve the performance of the proposed technique in terms of imperceptibility, capacity and robustness compared to existing techniques in the literature. In the future works, we will optimize and increase the quality of our algorithm by using the genetic algorithm.

REFERENCES

[1] Perednia DA, Allen A, "Telemedicine Technology and Clinical Applications".JAMA273(6):483-488,(1995) doi:10.1001/jama.1995.03520300057037.

[2] Zear, Aditi, Amit Kumar Singh, and Pardeep Kumar. "Multiple Watermarking for Healthcare Applications." Journal of Intelligent Systems (2016).

[3] Zear, Aditi, Amit Kumar Singh, and Pardeep Kumar. "A proposed secure multiple watermarking technique based on DWT, DCT and SVD for application in medicine." Multimedia Tools and Applications (2016): 1-20.

[4] Singh, Siddharth, Vivek Singh Rathore, and Rajiv Singh. "Hybrid NSCT domain multiple watermarking for medical images." Multimedia Tools and Applications (2016): 1-19.

[5] Muharemagic, Edin, and Borko Furht. "Survey of watermarking techniques and applications." Multimedia Watermarking Techniques and Applications 3 (2006): 91-130.

[6] Singh, Amit Kumar, Mayank Dave, and Anand Mohan. "Hybrid technique for robust and imperceptible multiple watermarking using medical images." Multimedia Tools and Applications 75.14 (2016): 8381-8401.

[7] V. Santhi, N. Rekha, S. Tharini, "A Hybrid Block Based Watermarking Algorithm using DWT-DCT-SVD Techniques for Color Images" Computing, Communication and Networking. International Conference on IEEE Xplore 18-20 Dec. 2008

[8] I.Assini, A.Badri, K.Safi "Adaptation of Different Techniques on Digital Image Watermarking in Medical Domain: A review", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Volume 5, Issue 12, December 2015.

[9] Takore, Tamirat Tagesse, P. Rajesh Kumar, and G. Lavanya Devi. "A modified blind image watermarking scheme based on DWT, DCT and SVD domain using GA to optimize robustness." Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on. IEEE, 2016.

[10] Sharma, Abhilasha, Amit Kumar Singh, and Satya Prakash Ghrera. "Robust and Secure Multiple Watermarking for Medical Images." Wireless Personal Communications (2016): 1-14

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Mitigating Economic Denial of Sustainability Attacks to Secure Cloud Computing Environments

Marwane Zekri[1], Said El Kafhali[2], Mohamed Hanini[3] and Noureddine Aboutabit[1]
[1]IPOSI laboratory, National School of Applied Sciences, Hassan 1st Univ, Settat, Morocco
[2]Computer, Networks, Mobility and Modeling laboratory
National School of Applied Sciences, Hassan 1st Univ, Settat, Morocco
[3]Computer, Networks, Mobility and Modeling laboratory
Faculty of Sciences and Technology, Hassan 1st Univ, Settat, Morocco
marwane.zekri@gmail.com, noure049@gmail.com; said.elkafhali@uhp.ac.ma;
haninimohamed@gmail.com

ABSTRACT

In cloud computing environment where the infrastructure is shared by millions of users, attackers have the opportunity to ensure more damage to the compromised resources. The main aim of such attacks is to saturate and overload the system network through a massive data packets size flooding toward a victim server and to block the service to customers. The Distributed Denial-of-service (DDoS) attack is considered one of the largest threats to the Quality of Service (QoS) of cloud services which is used to deny access for legitimate users of an online service. However, Economic Denial of Sustainability (EDoS) attack is a special breed of DDoS attack that attack exploits auto scaling feature of cloud. The Cloud Service Provider (CSP) scales the architecture automatically to serve those requests for which cloud consumer is charged. A consumer expects a sustainable profit after hosting his application on cloud. The attacker purpose is to guarantee the service unavailability and maximize the financial loss costs by increasing the cost and decreasing the profit. Hence, in this paper we propose a novel mitigation system against the EDoS attacks. Our system consists of source Checking, Counting, and Turing Test. The obtained simulation results show that our system works efficiently to mitigate the EDoS attack in cloud environment.

Keywords-component; Security, Cloud Computing, DDoS attack, EDoS attack, Mitigation

1 Introduction

Cloud computing refers to a type of Internet-based computing that provides shared pool of resources that can be rapidly provisioned on demand over the Internet [1], [2]. For instance networks, memory, computer processing, and user applications [3]. The cloud computing services can be categorized into three models: Software-as-a-service (SaaS), Platform-asa-Service (PaaS) and Infrastructure-as-a-Service (IaaS) [4], [5], and it can be deployed as private, public, community or hybrid cloud [2] (Figure 1). The benefits of cloud computing are several such as being cost savings, the scalability feature, reliability, maintenance, and mobile accessible. But besides all these benefits, cloud computing does come at the cost of increased security risks in which is at the moment one of the biggest challenges this technology is facing today, limiting the number of organizations willing to embrace it wholeheartedly [6]. DDoS [7] is

DOI: 10.14738/tmlai.54.3220
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3220

one type of aggressive attack, which causes serious impact on cloud servers, and according to the list of cloud security threat by the CSA (Cloud Security Alliance) as one of the most severe security threat for the cloud Computing and hinder it.

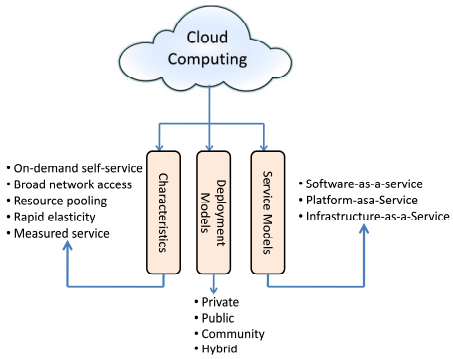


Figure. 1. Cloud computing characteristics, service and deployment models

DDoS attacks exhaust resources of a target [8], by flooding the target with spurious packets. The target could be a servers computing capability or network bandwidth such that the legal users could not access the services provided (Figure 2). EDoS attack is a special breed of DDoS attack that attack exploits auto scaling feature of cloud computing [9]. In EDoS attack, the goal in mind of the attacker is to manipulate service usage billing in orders of magnitude that could be disguised easily as legitimate use of the service [10]. This drives costs to unmanageable levels and maximizes the financial loss by decreasing the profit. Hence, an EDoS mitigation system becomes a technical and economical necessity to defense EDoS attacks in cloud system.

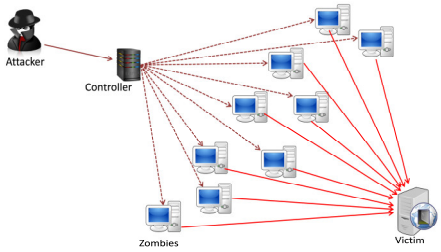


Figure. 2. A typical Architecture of DDoS/EDoS Attacks

To mitigate these risks, the cloud providers adopted many methods and techniques [11], [12] that can be used to prevent, detect and respond to attacks against the cloud. But in the other hand, as cloud strengthens its defense, the attackers find more creative and sophisticated ways targeting the cloud by using or creating loopholes in the cloud defense system. To keep up with this challenge and ensure the continuation and the existence of the cloud, we must improve, revise and upgrade our mitigation strategies. To this end, in this work we propose a novel mitigation technique against EDoS attack in cloud computing, to verify whether the requests coming from the users are from a legitimate person or generated by bots. In our mitigation technique we use a virtual firewall based on white, black and block lists filtering mechanism. The first request is forwarded to a Test Node. This node is responsible for updating the white, block and black lists based on the results of a text based verification challenge. The next requests coming from the bots will be blocked by a virtual firewall if their IP addresses source is found

in the black list. On the other hand, the requests coming from legitimate clients will be forwarded directly to the target cloud service since their IP addresses will be found in the white list. As a result, only the requests from legitimate clients will reach the Protected Target. If bots succeeded in bypassing the verification process it will be detected by the EDos Detecting Module which will put their IP Source in the block list.

Our main contributions in this paper can be summarized as follows:

- A text-based question which uses less bandwidth than an image based challenge.
- An algorithm for mitigation in case bots succeeded in bypassing the verification process.
- A simulation results that validate our proposed mitigation strategy.

The rest of the paper is organized as follows. First, in Section II, we discuss the related works. Section III presents the proposed architecture. The experimental setup including the simulation model and results are discussed in Section IV. Finally, the conclusion and the future work are presented in Section V.

2 Related work

Cloud computing EDoS Security has earned a lot of attention in the past few years, and a plethora of literature has been published on the subject. In this section we present published papers related to mitigate EDoS cloud computing system. EDoS Armor [13] protects E-commerce applications in cloud environments, with the assumption that an attacker does not follow regular workflow by purchasing the items instead idle surfing. EDoS Armor proposes muti-layered defense system that includes two modules:

- Admission control module: identifies whether the request is from legitimate client or bot by either image or cryptographic based challenge. Thus, it authenticates number of users in the system. Then, it allows limited number of valid clients to send the requests simultaneously through port hiding. This avoids over burdening.
- Congestion control module: this module handles the maximum resources available to legitimate clients. The client is categorized judging upon his browsing behavior using Decision tree algorithm J48 to classify the clients over the log file using parameters like Purchasing History, CPU Processing time, Session information, Resources Access Pattern.

Random Port Hopping Technique [14] is an In-client server communication scenario. Clients are using IP address and port number. Initially, server share cryptographic key to client for generating random port numbers. Authorized clients can use and determine the server port numbers and access the authorized information. Control measures are also unable to detect the application layer denial of service attacks. But attackers gain this information to target the high profile web servers to degrade their performance.

CloudWatch [15] is an auto scaling technique enabled by Amazon that reduces the effects of the EDoS attacks. CloudWatch is a web service that provides monitoring for cloud resources by which clients will be able to define boundaries that would limit the elasticity of their cloud platforms and thus reducing the effect of the EDoS attacks. In [16], the authors proposed a defending system by combining the filtering and detection mechanisms in order to mitigate the Denial of Service (DoS) attack in cloud data center environment. However, they presented an analytical model based on queueing model to evaluate the impact of flooding attack on cloud environment regarding service availability and QoS performance.

In-cloud scrubber service [17] is an on-demand service to mitigate network and application layers from EDoS. There are two modes based on resource depletion: normal and suspect. If the resource depletion

level goes beyond the limit and bandwidth traffic is also very high, then the service provider suspects high rate attack. The system switches to suspected mode and called scrubber Service which generate a puzzle to check legitimacy of the client. In [18], the authors proposed WebSOS to protect web servers by using a filtering mechanism that admits HTTP traffic from only trusted sources known to overlay nodes, to tell Computers and Humans Apart Legitimate clients have to first pass the Completely Automated Public Turing test (CAPTCHA) [19]. The proposed WebSOS presents the performance degradation due to non-direct routing, and also CAPTCHA is being challenged today by bots with the appearance of many algorithms that cracks CAPTCHAs such as Google Maps street address, which reading the algorithm with 99.8 percent accuracy. Furthermore, CAPTCHA based on image challenge [20], [21] needs quite the amount of bandwidth to transmit the CAPTCHA image to the users.

DDos-Shield [22] methodology for mitigating the EDoS in a cloud computing environment, has two main components a virtual firewalls (VF) witch based on a white and black lists filtering mechanism that hold IP addresses of the originating nodes, and the verifier cloud nodes which are represented by a pool of virtual machine nodes implemented based on the cloud infrastructure. The V-Nodes constitute a cloud-based overlay network. A V-Node has the capabilities to verify legitimate requests at the application level using graphic Turing tests, such as CAPTCHA. Another role of the V-Node is to update the lists used by the VF. This approach does not require hiding the location of the protected cloud service and using direct routing, on the other hand it requires the collaboration of an anti-spoofing methodology, and bots may bypass the CAPTCHA test.

In respect to these works, our proposed approach adds a filtering list to the firewall, precisely this list, named Block List, contents malicious IP sources that detected after being successfully blocked permanently or helps to relieve workload on the server.

3 Proposed Model

Figure 3 shows the proposed architecture of our approach for EDoS mitigating in the cloud computing environment. The firewall in our architecture is based on white, black and block lists filtering mechanism. The blacklist is used to hold those unauthenticated source IP addresses so that the firewall will drop the incoming packets originating from these IP addresses. The white list is used to track the authenticated source IP addresses so that the incoming traffic originating from these addresses will be allowed to pass the firewall towards the destined services. The Block list is a list that hold in a temporary manner the packets that are suspected to be malicious or with an abnormal traffic.

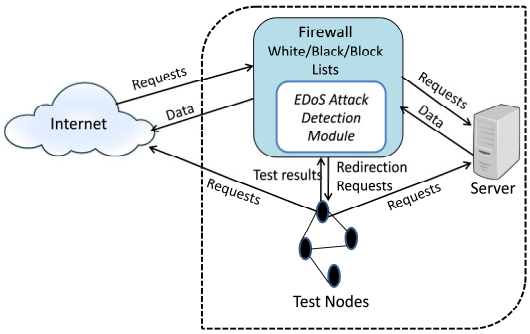


Figure. 3. Proposed mitigation architecture

The packet coming from the Internet will first arrive in virtual firewall. If the IP source exists in the White list, it will be forwarded to its destination. If it IP address source is not figuring in any of lists (Algorithm 1), the packet will be Forwarded to Turing test module by the test Nodes. The Test Nodes constitute a cloud-based overlay network which is represented by a pool of virtual machines. A Test node has the task of verifying the legitimacy of requests at the

application level using the Text based Turing tests and to update the black and white lists used by the virtual firewall (Algorithm 2).

Algorithm 1 Firewall Actions

$P \leftarrow$ Incoming Packet
$S_Ip \leftarrow$ Packet source
$D_Ip \leftarrow$ Packet destination
$BlackL \leftarrow$ BlackList
$WhiteL \leftarrow$ WhiteList
$BlockL \leftarrow$ BlockList
if $S_Ip \in WhiteL$ **then**
  Forward P **to** D_Ip
else if $S_Ip \in BlackL \parallel S_Ip \in BlockL$ **then**
  Drop **Packet** P
else
  Forward P to a Test-Node
end if

The Turing Test will challenge the requester by responding to a text-based question as for image challenge needs more bandwidth. To continue with the connection, the requester must succeed the challenge by returning the right answer in less than 3 attempts. As for The test Nodes they have the responsibility to update the lists based on the results of the challenge by marking IP address of the packet in the black, and white filtering lists of the firewall.

Algorithm 2 Test Node Actions

$P \leftarrow$ Incoming Packet
$S_Ip \leftarrow$ Packet source
$D_Ip \leftarrow$ Packet destination
$BlackL \leftarrow$ BlackList
$WhiteL \leftarrow$ WhiteList
Send to S_Ip a text Turing test
if the packet fails in the Turing test and attempts $\succ 3$ **then**
  $BlackL \leftarrow BlackL + S_Ip$
else if Turing test passes and attempts ≤ 3 **then**
  $WhiteL \leftarrow WhiteL + S_Ip$
end if

The role in Algorithm 3 is trying to find the possible malicious sources and add them into the Block or Black lists. Thus, even if the bots bypass the Request-challenge test or compromised by the attacker it will be detected, and these source IP address will be putted in either the black or block lists of the Virtual Firewall depending on the values of the defined thresholds given by the cloud customer. There are two Modes in the mitigation algorithm approach:

- The first mode named monitor mode (Passive): in this mode the Detection Module collects and records the traffic which will be used to detect the abnormal traffic and malicious packets. If the traffic data statistic is large than the maximum values of the thresholds defined by the cloud customer. The second mode will be triggered.

The second mode (Active): while the monitor mode keep running the second mode work with threshold values which can be enabled or fixed by the cloud customer as to not eventually gets the server overwhelmed. If the number of the connection times of an IP address exceeds Min-Num-Conn and is less than Max-Num-Conn, these IP source will be placed in the block list for a fixed short delay Roulette-Time-Block as for not exhaust the resource of the protected server. Once the delay expired the IP address source are removed from the block list. If the number of the connection times of an IP address is larger than Max-Num-Conn, this IP will be added to the black list (Algorithm 3).

Algorithm 3 Algorithm Mitigation Module

$S_Ip \leftarrow$ IP source
$Nb_conn \leftarrow$ Number Connections By S_Ip between two timestamp
$D_Ip \leftarrow$ Packet destination
$BlockL \leftarrow$ BlockList
$BlackL \leftarrow$ BlackList
while true **do**
  if Min-Num-Conn $\prec Nb_connInPeriod \leq$ Max-Num-Conn **then**
    $BlockL \leftarrow BlockL + S_Ip$
    $Delay\_level \leftarrow$ Roulette-Time-Block
  else if Roulette-Time-Block Is up **then**
    $BlockL \leftarrow BlockL - S_Ip$
  else
    $BlackL \leftarrow BlackL + S_Ip$
  end if
end while

4 Simulation Results and Discussion

In this section, we have carried out two different simulation scenarios. In the first, we have proposed that the cloud computing environment was not protected. There is any defense system to mitigate EDoS attack risk. In the second, we have implemented the proposed mitigation technique described above so as to decrease the impact of EDoS attack and to achieve a high level of security. Performance curves as those of the CPU utilization, mean response time, and the transmission and reception rate are plotted in the figures.

Figure 4 shows the obtained results regarding the response time. The results show that the response time increases considerably when not applying the mitigation technique. On the other hand, with the proposed mitigation technique, the response time corresponding to legitimate clients is approximately constant and low. The increasing of the response time when not applying the mitigation technique is due to the fact that legitimate requests suffer from more delays caused by the load of the attack traffic, which reaches the queues of the targeted cloud service. However, when using the proposed mitigation technique, the attacks requests will not affect the response time of the legitimate requests.

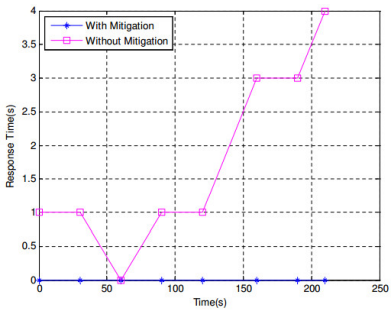


Figure. 4. Impact of EDoS attack on response time

Figure 5 shows the performance of the two scenarios in terms of the computing power utilization. The results show that, without a mitigation technique, when the arrival attack rate increases (by time), the computing power utilization increases indicating a high consumption of the computing resources by the attack requests. With the proposed mitigation technique, the computing power utilization is weakly affected due to the attack rate since the attack requests will not reach the protected cloud service. As a result, the computing power utilization with mitigation is not being changed significantly since the arrival rate of the legitimate requests is fixed.

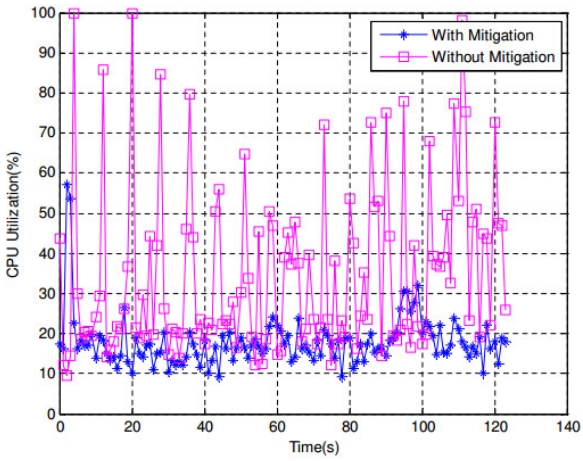


Figure. 5. Impact of EDoS attack on CPU utilization

Figure 6 shows the Transmission and Reception rates (TX/RX) in the two cases with and without applying mitigating. We remark that the approach with mitigation performs better for either Transmission or Reception rate parameter.

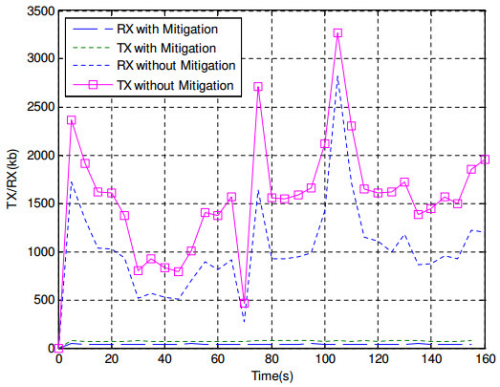


Figure. 6. Impact of EDoS attack on TX-RX rate

5 Conclusion and future work

EDoS attack seeks to disrupt the financial viability in cloud by exploiting utility pricing model by increasing the cost and lowering the profit. In this manuscript, we have proposed an EDoS mitigation system which consists of Source Checking, Counting and turning test. Algorithms that describe different operations of the proposed system are presented and discussed. The obtained results show that our approach reduces Computing Power Utilization and Response time, and performs better for Transmission and Reception rate in case of EDoS attack. As future work, we are planning to enhance our Attack Detection module and

integrate it with the proposed mitigation technique in order to rise the efficiency. We also aim to enhance our mitigating strategy to consider IP spoofing attacks.

REFERENCES

[1] Shawish and M. Salama, "Cloud computing: paradigms and technologies," in Inter-cooperative Collective Intelligence: Techniques and Applications. Springer, 2014, pp. 39–67.

[2] P. Mell and T. Grance, "The nist definition of cloud computing," 2011.

[3] M. Hanini and S. El Kafhali, "Cloud computing performance evaluation under dynamic resource utilization and traffic control," in ACM Second International conference on Big Data, Cloud and Applications (BDCA17). ACM, 2017.

[4] S. K. Garg, S. Versteeg, and R. Buyya, "A framework for ranking of cloud computing services," Future Generation Computer Systems, vol. 29, no. 4, pp. 1012–1023, 2013.

[5] S. El Kafhali and K. Salah, "Stochastic modelling and analysis of cloud computing data center," in 20th ICIN Conference Innovations in Clouds, Internet and Networks. IEEE, 2017, pp. 122–126.

[6] M. T. Khorshed, A. S. Ali, and S. A. Wasimi, "A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing," Future Generation computer systems, vol. 28, no. 6, pp. 833–851, 2012.

[7] O. Osanaiye, K.-K. R. Choo, and M. Dlodlo, "Distributed denial of service (ddos) resilience in cloud: review and conceptual cloud ddos mitigation framework," Journal of Network and Computer Applications, vol. 67, pp. 147–165, 2016.

[8] E. Alomari, S. Manickam, B. Gupta, M. Anbar, R. M. Saad, and S. Alsaleem, "A survey of botnet-based ddos flooding attacks of application layer: Detection and mitigation approaches," in Handbook of Research on Modern Cryptographic Solutions for Computer and Cyber Security. IGI Global, 2016, pp. 52–79.

[9] G. Somani, M. S. Gaur, D. Sanghi, and M. Conti, "Ddos attacks in cloud computing: collateral damage to non-targets," Computer Networks, vol. 109, pp. 157–171, 2016.

[10] M. Ficco and M. Rak, "Economic denial of sustainability mitigation in cloud computing," in Organizational Innovation and Change. Springer, 2016, pp. 229–238.

[11] N. Z. Bawany, J. A. Shamsi, and K. Salah, "Ddos attack detection and mitigation using sdn: Methods, practices, and solutions," Arabian Journal for Science and Engineering, vol. 42, no. 2, pp. 425–441, 2017.

[12] N. Agrawal and S. Tapaswi, "Defense schemes for variants of distributed denial-of-service (ddos) attacks in cloud computing: A survey," Information Security Journal: A Global Perspective, pp. 1–13, 2017.

[13] M. Masood, Z. Anwar, S. A. Raza, and M. A. Hur, "Edos armor: a cost effective economic denial of sustainability attack mitigation framework for e-commerce applications in cloud environments," in 2013 16th International Multi Topic Conference (INMIC), IEEE, 2013, pp. 37–42.

[14] R. P. Kumar, J. Babu, T. Gunasekhar, and S. B. Bhushan, "Mitigating application ddos attacks using random port hopping technique," International Journal of Emerging Research in Management &Technology, vol. 4, no. 1, pp. 1–4, 2015.

[15] A. CloudWatch, "Amazon cloudwatch," 2014.

[16] I. EL Mir, D. S. Kim, and A. Haqiq, "Towards a stochastic model for integrated detection and filtering of dos attacks in cloud environments," in ACM Second International conference on Big Data, Cloud and Applications (BDCA17). ACM, 2017.

[17] M. N. Kumar, P. Sujatha, V. Kalva, R. Nagori, A. K. Katukojwala, and M. Kumar, "Mitigating economic denial of sustainability (edos) in cloud computing using in-cloud scrubber service," in 2012 Fourth International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2012, pp. 535–539.

[18] A. Koduru, T. Neelakantam, and S. M. S. Bhanu, "Detection of economic denial of sustainability using time spent on a web page in cloud," in 2013 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM). IEEE, 2013, pp. 1–4.

[19] L. Von Ahn, M. Blum, N. J. Hopper, and J. Langford, "Captcha: Using hard ai problems for security," in International Conference on the Theory and Applications of Cryptographic Techniques. Springer, 2003, pp. 294–311.

[20] R. Datta, J. Li, and J. Z. Wang, "Imagination: a robust image-based captcha generation system," in Proceedings of the 13th annual ACM international conference on Multimedia. ACM, 2005, pp. 331–334.

[21] A. Gupta, A. Jain, A. Raj, and A. Jain, "sequenced tagged captcha: generation and its analysis," in IEEE International Advance Computing Conference, 2009. IACC 2009. IEEE, 2009, pp. 1286–1291.

[22] M. H. Sqalli, F. Al-Haidari, and K. Salah, "Edos-shield-a two-steps mitigation technique against edos attacks in cloud computing," in 2011 Fourth IEEE International Conference on Utility and Cloud Computing (UCC). IEEE, 2011, pp. 49–56.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Architecture of A Semantic Annotation of Handwritten Documents Based on the Ontology "OMOS"

[1]Mint Aboubekrine Aiche, [2]Mohamed Lamine Diakité, [3]Kamal Eddine EL Kadiri, [4]Youness Tabii
[1,3,4]LIROSA Laboratory; Systems Research Uni, ENSA , Tetouan, Abdelmalek Essaadi University, Morocco
[2]URDNI Research Unit, Faculty of Science and Technology, University of Nouakchott Alaasriya
ne3ma.20@gmail.com, diakite_medlamine@yahoo.com, elkadiri@uae.ma, youness.tabii@gmail.com

ABSTRACT

In this work we propose to present an application that supports the representation of manuscript documents according to an ontological approach. The implementation of this application makes it possible to annotate semantically these manuscripts according to the ontology "OMOS" [1]. The semantic annotation proposed here is an annotation based on an ontology of the "OMOS" manuscripts of generation and use of specific metadata targeted to allow new methods of access to information and to extend existing ones. The proposed annotation is based on the understanding that the named entities (the author, date 'link ... etc.) mentioned in the documents constitute an important part of their semantics. Finally, this application gives us an annotation file associated with each document.

Keywords: Semantic annotation, Ancient Manuscript, Ontology.

1 Introduction

As part of the safeguarding of the world's cultural heritage, digitization campaigns of ancient manuscript documents that constitute the cultural heritage of nations are becoming increasingly common. This digitization campaign generated a large number of digitized resources with an invaluable amount of knowledge. Likewise, a large number of ancient works and documents from all over the world are kept in archives which are threatened with disappearance due to moisture, acidity of paper, etc. It is therefore important to preserve this heritage, to make it accessible to everyone and especially with easy access. Digitization is the solution adopted, but it only provides images of documents, which is not always sufficient. Indeed, it is often necessary to access the contents of digitized documents and to exploit them possibly. This is the object of semantic annotation of these ancient manuscripts. To ensure the efficiency and reusability of metadata, we base ourselves on the OMOS ontology [1], it is based on specific knowledge of African handwritten documents, rather than being indifferent to any ontological commitment and any knowledge, but a vast knowledge base of the descriptive entities is maintained on the Arabic manuscripts. To carry out this work there are several state-of-the-art documents dealing with particular documents or documents of general application [3]. Here we present the Semantic annotation technique proposed. Finally, in this work we will describe the design of (ASDM) realized for the modeling of the semantic annotation based on the OMOS ontology. Our vision is that fully semi-automated methods for semantic annotation should be sought and developed. To do this, the necessary design and management issues must be extracted and annotated, and the necessary additional resources and

DOI: 10.14738/tmlai.54.3221
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3221

infrastructure must be provided. To ensure wide acceptance and proper use of semantic annotation systems their tasks must be clearly defined and their performance properly assessed and communicated.

2 Related Work

Several works on the semantic annotation exist in the literature. On the research side, several works exist for the semantic annotation among these works we mention some of them according to the field of research [3, 4, 5]. In another part of the research, a few works on ancient Arabic manuscripts exist [1], [6].First in the semantic web, semantic annotation is defined as the process that determines the interpretation of a document by associating it with formal and explicit semantics. This interpretation is commonly expressed in ontological terms when it is a question of associating a semantic type with the names of the entities mentioned in the text [7,8].The image format in which these documents are saved makes it difficult to be able to exploit their content satisfactorily, whereas current semantic annotation work exploits information retrieval techniques to associate document strings with of concepts of ontology [4].This annotation can be manual, automatic or semi-automatictexts. In return, we do not find an interested approach to the annotation of handwritten documents. The idea is, once digitized manuscripts, to facilitate their access to researchers and other users through a system and a thematic organization. As this annotation is expressed in ontological terms, we find that [1] OMOS ontology is the first one for Arabic heritage manuscripts. The objective is to find a way that will allow the needs of the users to be better interpreted and thus allow a better access to the content. To do this, we proposed a modeling of knowledge on manuscripts through the ontology "OMOS". Its objective is to implement an application for the annotation of ancient manuscripts. This application simplifies and facilitates access to these documents. Indeed, it exploits the descriptions associated with the image to construct semantic annotations

3 Document Semantic Annotation Technique

3.1 Description of a Document

Contrary to the classical representation of a document that only takes into account the textual content [3] and where the information is presented in a flat manner without any treatment, the structural presentation of a manuscript must enable us to identify At least two different levels of knowledge that characterize manuscripts.

The first level corresponds to the descriptive and situational knowledge that applies to the exploitation of the manuscript whatever its content.

The second level is linked to knowledge coming directly from the contents of the manuscript. Obviously, this type of knowledge is at the same time more difficult to obtain, [4] much more diversified and less limited than the previous level of description.

We started with the first level.

The following figure (1) presents the descriptive knowledge that could be used to properly present the manuscripts. This descriptive knowledge is available in the catalog of the IMRS library enriched by the suggestions obtained by the experts.

Information		About Resource	Type
Author	المؤلف	Manuscript	Concept
Title	العنوان	Manuscript/Copy/Copy with comments	Attribute
Subject	الموضوع	Manuscript/Copy/Copy with comments	Concept
Copyist	الناسخ	Copy	Concept
Write mode	الخطّ	Manuscript/Copy/Copy with comments	Attribute
Ink color	الحبر	Manuscript/Copy/Copy with comments	Attribute
Page surface	قياس الصفحة	Manuscript/Copy/Copy with comments	Attribute
Page number	الصفحات	Manuscript/Copy/Copy with comments	Attribute
Written surface	مساحة النص	Manuscript/Copy/Copy with comments	Attribute
Line number	الأسطر	Manuscript/Copy/Copy with comments	Attribute
Entire or not	النص تام ؟	Manuscript/Copy/Copy with comments	Attribute
Year[16] of publication	تاريخ النشر	Manuscript/Copy/Copy with comments	Attribute
Incipit[17]	البداية	Manuscript/Copy/Copy with comments	Attribute
Explicit[18]	النهاية	Manuscript/Copy/Copy with comments	Attribute
Exegetes		Copy with comments	Concept
Birth year / Death year		Author/Copyist/ Exegetes	Attribute
Birth Place		Author/Copyist/ Exegetes	Concept
Place of origin		Author/Copyist/ Exegetes	
Library		Manuscript/Copy/ Copy with comments	Concept
Location		Library / Private Library / Public Library	Concept
Language[19]		Manuscript/ Copy/Copy with comments	Concept
Field		Manuscript/ Copy/Copy with comments	Concept
Named time period		Manuscript/ Copy/Copy with comments	Concept

Figure 1: descriptive knowledge of manuscripts

3.2 Presentation of OMOS ontology

Ontologies traditionally play an important role in data integration processes [9],[10], which is also a key property when dealing with semantically heterogeneous sources such as handwritten documents. We present OMOS, an ontology describing West-Saharan manuscripts in Africa. The first version of an ontology on the manuscripts was realized "Semi-automatic construction of an ontology on Western Saharan manuscripts" we will try to use this ontology to semantically annotate the manuscripts efficiently. Our approach is based on the use of an OWL ontology that deals with the semantic web domain. This ontology is called OMOS, it is developed by the LABORATORY of INFORMATIQUE LI of the University François Rabelais of Tours in FRANCE. Is ontology contains hierarchies of concepts focusing on the domain of the semantic Web. Figure 2 presents the hierarchy of this ontology.

Figure 2: presents a hierarchical part of this ontology

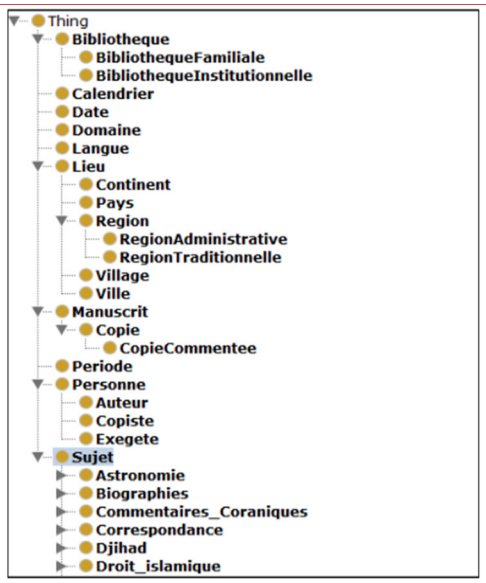


Figure 3: presents some of this ontology

3.3 Semantic annotation proposed

The semantic annotation of web resources is a way of moving from the current web to the vision of the future Web envisaged by Tim Berners-Lee. Indeed, it represents a process which aims to formalize semantic interpretations extracted from the document. The proposed technical process is a design of an application, which takes as an input step another approach. This application is a process that allows us to realize our object in order to offer us an annotated document. The latter determines a new document associated with semantic data.

4 Modeling Of A Sementicannotation

In this section we will present the architecture of our application and the implementation work we did, which consisted firstly of the edition of our ontology under the OWL language, followed by its exploitation in a semantic annotation application, which was developed in Java.

4.1 Semantic annotation architecture

In order to implement this application we propose a design presenting the structure of our work. In all cases, the semi-automatic annotation method will be adopted. The application will request the handwritten document of image format plus the information provided on this document andby monitoring the application can enrich these descriptions based on its own knowledge modeled in the form of the ontology of the domain. This will be discussed in the next section (see Figure 4).

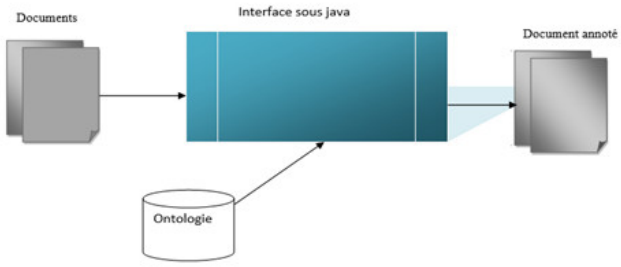


Figure 4: The architecture proposed for the application

4.2 Implementation

The implementation of our approach proposes an application implemented in JAVA and using the Jena library to exploit the different relationships. This JENA library contains a Java API for the manipulation of RDF / RDFS and OWL files, for the diffusion of ontology and the management of queries.

5 Result

These methods mainly use two concepts, namely the ontological approach to semantically apply and model the metadata, and the annotation application that allows to associate interpretations with the handwritten documents and facilitates the communication between its different Components. Another interesting line of research to be devoted to it is the identification and the extraction of the information of the manuscript like the name of the author, the copies and date of creation and the place etc. This information can then be semantically annotated according to an OMOS ontology-based application. This allows you to search annotated information.

6 Conclusion

We chose this methodology as it allowed to formalize and generate all the explicit and implicit knowledge on the manuscript and to annotate it semantically according to the OMOS ontology.

As research perspectives, we have identified this theme that it would be interesting to explore. The proposed application may ultimately serve as a basis for research in manuscripts.

The different concepts defined there will serve as a semantic repository to better interpret user queries.

REFERENCES

[1] Mohamed Lamine DiakitéBéatrice Bouchou Markhoff, « Construction semi-automatique d'une ontologie sur des manuscrits ouest sahariens » Communication dans un congrès IC2015, Jun 2015, Rennes, France. AFIA.

[2] Yue Ma, Laurent Audibert, Adeline Nazarenko « Ontologies étendues pour l'annotation sémantique » 20es Journ_ees Francophones d'Ing_enierie des Connaissances, May 2009, Hammamet, Tunisie, Tunisie. pp.205-216, 2009. <hal-00378594>

[3] A. Guissé, F. Lévy, A. Nazarenko, and S. Szulman «Annotation sémantique pour l'indexation de règles métiers » Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA 2009), page (electronic medium). Université Paul Sabatier - Toulouse, (November 2009).

[4] UREN V., CIMIANO P., IRIA J., HANDSCHUH S., VARGAS-VERA M., MOTTA E. & CIRAVEGNA F. (2006). Semantic annotation for knowledge management.

[5] Teyeb, M. Torjmen, N. Hernandez, O. Haemmerlé, and M. Jemaa. « Vers une annotation sémantique des images web fondée sur des patrons RDF ». CORIA, page 285 300. ARIA, (2015).

[6] Bertrand Coüasnon, Jean Camillerapp, « Accès par le contenu aux documents manuscrits d'archives numérisés », Document numérique 2003/3 (Vol. 7), p. 61-84 DOI 10.3166/dn.7.3-4.61-84.

[7] KIRYAKOV A., POPOV B., TERZIEV I., MANOV D. & OGNYANOFF D. (2004). Semantic annotation, indexing, and retrieval. J. Web Sem., 2(1), 49–79.

[8] Nguyen M., « Vers une plate-forme d'annotations sémantiques automatiques à partir de documents multimédias », 2007.

[9] H.Wache, T.Vögele, U.Visser, H.Stuckenschmidt, G.Schuster, H.Neumann, and S.Hübner, Ontology- Based Integration of Information – A Survey of Existing Approaches. IJCAI Workshop on Ontologies and Informations Sharing, pp 108–117, 2001.

[10] Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov, Damyan Ognyanoff « Semantic annotation, indexing, and retrieval » Web Semantics: Science, Services and Agents on the World Wide Web, Volume 2, Issue 1, Pages 49-79.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Use of the MOOC for a Change of Attitude Towards Science and its Teaching

Rajaa Aoulad Moussa, Mourad Madrane, Rachid Janati-Idrissi, Mohammed Laafou
Interdisciplinary laboratory of Research in Pedagogic Engineering, ENS-Tétouan, Morocco
rajoua78@yahoo.fr

ABSTRACT

This work describes conceptual changes in science and its teaching; Observed among 200 primary school teachers after the application of a didactic proposal (on-line training).

We used a platform as training and learning space to apply the didactic proposal.

Attitudes were assessed by means of an instrument, applied before and after the proposal. The instrument included the three traditional components of attitude (cognitive, affective and active) across different scales, semantic differential and forced option reagents. The results of the preliminary survey, carried out before the didactic intervention, revealed that primary school teachers have notions, emotions and actions, which translate into unfavorable attitudes towards natural sciences, which directly affect their teaching practice.

After the application of online training, it was obvious both qualitatively and quantitatively a positive change in the attitudes of teachers towards science and its teaching.

Keywords- platform, science; Change of attitude; Science teachers; online training; primary education

1 Introduction

One of the missions of the teaching of sciences is to make it possible to the future citizen to acquire a minimum of scientific culture. However as shown by the results of several studies and investigations (PNEA 2008) [1]and (TIMSS 2003, 2007, 2011) this task appears difficult. A concern is raised in Morocco following the disaffection of the course of scientific training. It is related to the fear of lacking scientific executives to ensure the economic development of the country and citizens equipped with a basic scientific culture, as well as to the will to improve the situation in this field. These concerns do not go back to today since they appear already in the speeches of the scientists of the Twenties and in the reports of research by 1948 [2].

The research results in connection with the attitudes of the pupils towards sciences show that they constitute a good predictor of the intention to engage in scientific studies, alone [3], or associated with other elements relating to social environment or to personal experience[4].

In this research, we will focus on the attitudes, because of their close and undeniable relationship with the teaching of sciences. For several years, it has been recognized that the emotional variables are as important as the cognitive variables to influence the training. The emotional variables are regarded as

DOI: 10.14738/tmlai.54.3223
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3223

important not only for their influence on the use, but because they are themselves the very relevant results of training [5].

For this reason the goal of this study was to guide a favorable change of attitude towards sciences and their teaching among primary school teachers by an educational proposal based on creative and entertaining activities.

In this study, we suggest an alternative solution to cure these problems through the development of a teacher training using an e-learning platform, which would be much more advantageous than the one we would provide within the framework of classroom training [6]. The reasons that justify this choice are of economic and didactic order. Compared to this last point, such teaching would make it possible for each teacher to learn from his knowledge of the subject of study, which is difficult, not to say impossible, in classroom teaching. [7].

2 Foundation and Development of On-line Training

One of the main problems of the teaching of sciences in Morocco is the difficulty which the teachers have to conceive of the strategies of suitable teaching to help their pupils develop a positive attitude towards science and as results they adapt the scientific knowledge.

The conceptions of the pupils and the teachers regarding science were the object of many works[8][9][10][11] These attitudes, analyzed in relation to epistemological references, are in most cases incoherent, kinds of several epistemological models, around an identified reason as holding mainly of a model which sends back to empiricism as to the status of knowledge, to realism as to their relation with the objects they talk about, to positivism as to the existence of a standard and a-historical approach to ensure its validity and to induction as regards the development process.

A research Moroccan provided by A.Moussa& al presented to the CIMQUSEF12 in 2016, shows that the majority of the primary school teachers adopt unfavorable attitudes regarding science, because of, among other things, the lack of mastery of the scientific contents and the lack of knowledge about practical and experimental activities [12].

According to Duit and Treagust, in order to go beyond this more traditional teaching, the conceptual change can be a research track“Research has shown that conceptual change informed teaching usually is superior to more traditional means of teaching. Hence, conceptual change may still be a powerful frame for improving science teachingand learning.”[13]

Primary school teachers must thus deeply know the scientific contents, which they teach, recognize their attitudes, recognize how to teach sciences and develop aptitudes and necessary competences to promote the positive attitudes among their pupils. This leaves the teachers in front of the challenge to face their own practice in order to try to transform their teaching strategies, an extremely difficult task, because of other factors, the lack of an initial adequate training[13].

An e-learning was conceived the goal of which is to direct a positive change of attitude towards sciences and its teaching. The training is founded on a theoretical and methodological reflection on the approach of traditional teaching practiced in the majority of primary school classrooms, to try to transform it into a constructive process where teachers can exert their creativity and where pupils build their scientific knowledge in an interesting and amusing way[6].

For the development of this training we took account of the orientations to elaborate international strategies and programs, in which the value of nature are recognized; facilitates the interaction between elements of primary education program and the community; includes practical activities which facilitate the conception of science as a dynamic process and in construction related to reality; presentation of the contents as unfinished and significant for the students and to develop a critical and reflexive awareness. These global directives are founded, from a constructivist point of view, in which the attitudes of the teachers and the pupils play a very important part in the training process [14].

Moreover, from this constructivist perspective, to explain the strategies used in the teaching of natural sciences, it is necessary to define first the conditions of the subject, the learning process, the contents and the objectives, the teaching equipment, the of evaluation ends and the role of the teacher.

The subject covered in this training was that of "nutrition" it is part of the official curriculum of Moroccan primary education. We consider it important to propose this subject for some reasons. Firstly, the importance for the children to acquire the fundamental knowledge to understand the natural phenomena, mainly those in relation with health preservation. Secondly, the importance of an adequate nutrition. It is present along the six degrees of primary education and thirdly, because it is a subject which teachers fail to present in an interesting way to their pupils[1].

The activities were designed as cognitive, emotional and active strategies to direct teachers to a positive change of attitude towards sciences and at the same time to help them understand in a profound and amusing way the contents that they will learn. The didactic approach was realized through activities that tried to facilitate, for the teachers, the management and the teaching of some contents on nutrition, in a way that they were translated into significant learning for the pupils. Through this e-learning teachers were expected to:

a) Be able to link the science contents to the daily reality their pupils.
b) Acquire the didactic component of scientific contents.
c) Promote their creativity to develop didactic strategies feasible and realizable in class.
d) Encourage the development of competences that support a skill so much in the teaching of science as in other subjects.

3 Reason and Advantage Of The Use Of Mooc For The Application Of Teacher Training

During their initial training, primary education teachers acquire a vast knowledge which is characterized by what is called "knowing the subject to be taught". Learning these contents is a requirement although nonsufficient, i.e. it does not guarantee an automatic transfer to the classroom practice in [15][12].

To correct the gaps in the initial primary education teachers' training, it is necessary to provide them with didactic materials that make it possible to work, in class, with the learned contents concerning the nature of science and that enable them to integrate in a single structure varied knowledge from the subject to be taught

The national Charter of education and training stipulates that any teacher has to profit obligatorily from a 30 hours training a year, and it is only in the 2009-2010 school year that this principle was applied for the first time to all primary education teachers (5 days of training at a rate of 6 hours a day on the

competence-based approach and the pedagogy of integration) [7]. The application of this principle raised other issues relating to the trainers' qualification, the material conditions of the training occurrence and the securing of the pupils' school time during the teachers' absence due to the training.

From this point of view, the development of a model of distance training would be much more advantageous than that which could be provided within the framework of a classroomtraining [6]. The reasons which justify this choice are of economic and didactic nature. Compared to this last point, such training would make it possible for each teacher to learn how to start from his knowledge on the subject of study, which is difficult, not to say impossible, in a classroom training [7].

3.1 Choice of platform

There exists a variety of e-Learning platforms – which diffuse massive, open and free online courses (MOOC) - available on the Web [16]. Each one of them has different characteristics that make it flexible to adapt to various training contexts. The choice of eFront learning platform was taken for the following reasons [17]:

- It is a complete and effective platform.
- It is easy to use
- It is free software: One can personalize it according to the needs for the training and the learners.
- It is compatible with several operating systems
- It is of last generation.
- It is an intelligent platform: The trainers can set access rules to the resources (obligation to read, to get certain results…). The trainer can thus easily define a teaching plan of training.

3.2 Choice of the training mode

According to Walkiers and Praetere, the collaborative training is any learning activity carried out by a group of learners having a common goal, being, each of them, a source of information, of motivation, of interaction, and of mutual aid and profiting from each other's contributions, of the group synergy and the assistance of a trainer facilitating individual and collective leanings.[18]

The online collaborative training encourages reflection, resource sharing, autonomy, the critical spirit and of synthesis. It is a method in agreement with the proposals of this training, since it is based on the following concepts:

- The communicational aspect of TIC can support a collaborative work by offering tools of communication, of creation and access to a large variety of resources [17]
- The subject of knowledge is autonomous;
- Cooperation constitutes the fundamental activity for expansion, knowledge structuring and resolution of identified problems;

Our distance training is based on an online training platform that offers the teachers the opportunity to publish their learners' cognitive problems and to try to analyze them by exploiting the various communications tools of the platform used.

The intention of the collaborative online training is to encourage an active training and to develop co-operative work among teachers.

4 Methodology

We are interested in comparing the teachers' attitudes before and after being subjected to a didactic proposal online. Participants were not randomly selected; the accent of research was quasi-experimental, by means of a questionnaire, applied before and after the proposal (pre-test - post-test). The instrument included the three traditional components of the attitude: cognitive, emotional and active through various Likert scales: semantic differential and reactive of forced option [19].

4.1 Sample

A sample of 200 teachers working at the public primary schools, of the provincial direction of national education and of vocational training of Tétouan.

These teachers have an average seniority of approximately 14 years. 65% are women. All participants have had an initial training in teachers training centers but the duration of their training varies. 68% had only one year training while others had two years.

4.2 Questionnaire in connection with the attitudes

To evaluate the possible changes in the attitudes of the teachers, a questionnaire was provided to them on the attitudes related to science before and after the proposal, which includes the three traditional components of attitude (cognitive, emotional trend and active) by Likert's scales, semantic differential and reactive forced choice. This instrument is composed of several parts: the first part comprises teachers' biographical and university data (age, sex, years of experience and the number of students). The second part includes questions about their attitudes in regards science, its teaching and about the scientific and experimental activities (emotional component). Thanks to semantic differential scales; in the third section teachers were questioned on their knowledge and beliefs related to sciences, their teaching, the scientific research and the characteristics of scientists (cognitive component). In the fourth section, the questions related to the activities of scientific research (component cognitive). In the fifth part, the questions were about teachers' preferences concerning science activities (component of trend). The last part was about opinions in connection with science and the scientific and technological research.

4.3 Procedures of analysis

The obtained results were organized and classified in categories to analyze them. For profound sessions we carried out a general qualitative analysis of the contents of information provided by the teachers, by establishing categories and units of analysis. Categories: a) topic, referring to the covered issues like science and the production of scientific knowledge, its relationship to technology; a teaching and a learning of science and an improvement of teaching strategies. - b) direction, relative to the issue which was covered, in a favorable or unfavorable way and c) value, relative to the importance the teachers granted the covered issue. Units of analysis, being 1) words, science, technology, didactics, contents - between the others - and 2) the subject, science and its teaching.

The data of the pre-test and post-test were analyzed in a global way by obtaining the descriptive statistics - an average, a standard maximal and minimal error of the average, of values - and the answer percentages, with an objective to describe teachers' changes and trends of attitudes .

We applied Statistical Package for the Social Sciences, SPSS version 12 in order to see whether there were significant differences between the pre-test and the post test, taking into account the size of the effect observed, statistical on that which carries out the interpretation of the size of the observed differences. In other words, between the teachers' attitudes before and after the proposal; the values and the averages of greater value indicated a positive or favorable direction of attitudes and the averages of less value a negative or unfavorable direction of attitudes.

5 Results and Discussion

5.1 Teachers' comments and opinions related to sciences

In the first part of the training, by means of depth sessions, evoked teachers' reflection and discussion through detailed questions in order to know the comments and opinions of the primary school teachers participating in this study related to science.

Questions 1 and 2 were conceived with an aim to make teachers explain the role that science plays in our society and what are, in their opinions, its objectives and characteristics.

The majority of answers (66.67%) affirm that science tries to give answers to the phenomena or the problems of nature. Only 16.5% explicitly mentioned the social context where the scientific activity is developed and only 16.66% underlined the specific character of scientific methodology. But, the important aspects of epistemology of contemporary science like creativity, imagination or personal subjectivity and the influence of research programs on the work of the scientists were mentioned by no teacher.

As for the characteristics and the objectives of science, they stress that the scientific knowledge is obtained by means of empirical evidence (50%) allot it values such as "objectivity" and "exactitude".

It should be stressed that almost 33.3% of the answers present unrelated and indecipherable explanations. This is why we can as well conclude that there exists a preliminary lack of reflection on the objectives of science and its social function.

Questions 3 and 4 were conceived so that the students explain the role of the experiment in the scientific activity. Question 3 is a direct one that aims at extracting their definitions of the characteristics of an experiment and question 4 interrogates them on the role that experiments play in the development of a scientific research.

66.67% of the teachers define the experiment as a try to empirically check a declaration or a theory, although they do not specify how the theory to check appeared. More than a third of the answers explicitly state that it is necessary firstly to develop the assumptions, which will subsequently be contrasted by means of experiments. In no case they include the definition of experiment nor explicitly state that it is part of a more total and complex research. However, practically all the students consider that the scientific knowledge needs experiments to be developed.

In coherence with the previous results, a large number of explanations justifies the importance of experiment as empirical verification of theories, not indicating any other rate of validation of these, its

predictability, universality and coherence with the theoretical framework. A third of the respondents considers that starting from experimentation we produce explanations or theories, in a clear empirical-inductive conception of science.

The purpose of questions 5 and 6 is to examine what teachers think of the role of theories in scientific knowledge. Question 5 is a direct one about the reliability of scientific theories and question 6 wonders about the difference between empirical evidence and explanatory theories.

With regard to what teachers think of the role that theories play in scientific knowledge, we notice that almost 66.67% confer a very high degree of certainty to theories, while only 16.66% indicate an average certainty and nobody allots them a low degree of certainty. However, practically all the justifications with this high degree of confidence to the scientific theory refer to the experiments they use to demonstrate it, as well as to the current technological devices. In the explanations, the difference between the experimental facts and the theoretical concepts conceived to explain them is not mentioned.

These results confirm the preliminary investigations which they report that an empiricist designs in connection with sciences still prevail among teachers, and in which the scientific knowledge is regarded as isolated truths and the training of those must be invariable [20]. On the production of the scientific knowledge. A good part of them (75%), mentioned that this kind of knowledge was generated in a laboratory while following each step of the scientific method, therefore. It is important on the one hand, to get rid of the idea saying that the production of scientific knowledge only happens and exists in the laboratory (Fourez, 1992). On the other hand, to eradicate this rigid vision they have with regard to the scientific method, which is considered a series of steps or of stages that must be followed in strict order and in a mechanical way. This is why, a teaching training is very important; that which includes an adequate comprehension of the nature of scientific knowledge (Abd-El-Khalick and Lederman 2000).

5.2 Attitudes towards biology and teaching

In the second part of the proposal, we worked with primary school teachers on the subject of “nutrition” through cheerful activities. Generally, the results showed favorable changes of attitude. The tables which are presented then show the results by component of attitude, obtained in the pre-test and post-test of various scales included in the instrument applied.

With regard to the emotional component of the attitude, tables 2 and 3 show descriptive statistics of the emotions expressed by the teachers towards science and its teaching obtained in the pre-test and post-test of the total items of the semantic differential referring to the emotions that science (table 1) and its teaching (table 2) produce among the questioned teachers.

Table 1: Descriptive Statistics of the Feelings which Sciences Produce in the Primary’s Schools Teachers.

	Average	NR	Deviation Standard	Error Standard	Min	Max
Pre-test	482.7155	200	7.80816	1.84040	25	54
Post test	773.4555	200	3.97254	.93634	39	56

Table 2: Descriptive Statistics of the Feelings which Science Produces among Primary School Teachers

	Average	NR	Deviation Standard	Error Standard	Mini	Maxi
Pre-test	570.9877	200	9.6779	2.2811	35	67
Post test	721.6044	200	3.0190	.7115	60	70

In these tables we notice that the averages of the post-test in these scales of attitude were larger than those of the pre-test. The average trend of emotions in the pre-test proved unfavorable, while in the post-test it tended to be very favorable.

With regard to the beliefs and the knowledge expressed by the teachers - of this cognitive sample-component of the attitude- in table 6 we shows the descriptive statistics obtained in pre-test and post-test through a standard scale Likert type. In this table, we notice an increase in the total average in post-test, which indicates a trend of change in the beliefs after the application of the proposal.

Table 10: Descriptive Statistics of the Beliefs and Knowledge in Keeping with Science and its Teaching Among Primary School Teachers.

	Average	NR	Deviation Standard	Error Standard	Min	Max
Pré-test	641.3577	200	6.26668	1.47707	41	74
Post test	735.8022	200	3.67112	.86529	61	67

To obtain information on the active component or of the trend to action of the attitude, we asked teachers to express their preference to succeed in certain activities in relation and without relation with science. The results are shown in tables 4 and 5.

Table 11: Descriptive Statistics of Teachers' Preferences for Science-Related Activities before and After Training.

	Average	NR	Deviation Standard	Error Standard	Min	Max
Pre-test	91.3577	200	4.03741	.95163	0	14
Post test	124.6911	200	3.11648	.73456	5.00	15

The data of table 4 shows an increase of the average in post-test. This means that teachers expressed a greater preference to bring to succeed in activities related to science after the application of the didactic proposal.

In table 5 we notice the test results of t testing and the size of the effect with regard to the active component. The change in the average of the post-test is significant with a $p<0.01$ and the size of the effect proved to be greater than one.

Table 12: Changes in the Preferences of Primary School Teachers (*Significant At P <0.01).

Preferences for :	Average Per-test	Average Post-test	Size effect	Error standard	t	Sig
Conduct activities related to science	91.3511	124.6911	1.0237	.69074	4.343	.000*

The fact that teachers expressed, before the implementation of the training, their preference to carry out non-scientific activities could be due to the difficulty in understanding and explaining the scientific knowledge and to the lack of clarity of scientific activities (Seam, 2002). After having applied the activities of the e-learning, teachers expressed a greater preference for the realization of science-related activities. We also noticed, on the part of the participating teachers, more enthusiasm, which enables us to suggest, on the basis applied by Perkes[21] – teachers who feel well prepared and who have self-confidence teach science better - that they have become more self-confident to carry out simple practical activities and to teach the contents of scientific programs.

By analyzing the three components of the attitude, we find that the attitudes of the teachers of this study were unfavorable towards science - as they shown by the results of the pre-test of the three components of the attitude. When teachers had taken part in the training, they showed significant changes and they showed us a more favorable attitude towards science and its teaching. The analysis of the results makes it possible to consolidate the assertion that to develop good school science and to get to significant knowledge one requires a positive attitude, an interest and a will.

After the application of the e-learning, we notice among the teachers a greater openness, a greater self-confidence. However, it is important to admit that the teachers constantly tried to provide their pupils with significant teaching and reflective thought but they do not always succeed. For that, we want to explain that with this proposal we try to direct only a favorable change of attitude and through that, to provide the teacher, in the simplest way, with some elements that help him in his task as a facilitator of learning.

6 Conclusions

The obtained results of this study show that primary school teachers are very motivated to profit from training. The suggested solution is based on the distance training that gives a very great flexibility to the teachers on the level of time and space.

The results also showed a favorable change of attitude towards science and its teaching. It is important to mention that for this change to be permanent we have to continue to innovate and to provide teachers with a continuous training.

This study showed that it is possible to change negative attitudes into positive attitudes towards science and its teaching, by means of a didactic proposal based on a reflection on the teaching practices in class. However, it is important to mention that this training was applied in a short time to a population of 200 teachers belonging to only one delegation. This is why it has to be taken into account that it is not possible to generalize these results. The data obtained from this research reveal that teachers need to be provided with strategies in order to help them to better direct the construction of knowledge among their pupils.

The results of this study enable us to suggest that to obtain a positive change of attitude towards sciences and its teaching, it is necessary firstly that the teachers become aware of their own attitudes and thus, will subsequently be able to develop favorable attitudes through self-confidence and reflection on their own teaching practices.

In addition, the results of this research revealed important educational implications. It is suggested that besides laying out positive attitudes towards sciences the professors:

b) Show to their pupils the utility of scientific knowledge so that they are interested in science and get to a significant learning.

c) Integrated to other continuous online training programs to be able to develop their innovating strategies of teaching-learning of sciences.

d) Take into account the social and environmental context to build significant processes of teaching-learning.

Finally, in this research we consider that if we are interested in the development of a high quality teaching of science at the primary school, then we must start with a change of attitude.

REFERENCES

[1] Superior Council of Education, "Analytical Report," MOROCCO, 2009.

[2] J BENNET, "The development and use of an instrument to assess students' attitude to study of chemistry," International Journal of Science Education, pp. 833-845, 2001.

[3] Noman REID and Elena A SKYABINA, "Attitudes towards Physics," Research in Science & Technological Education, vol. 20, pp. 67-81, 2002.

[4] Frank E CRAWLEY and Annette E COE , "Determinants ofmiddle school students' intention to enrol in a high school science : an application of the theory of reasoned action," Journal of research in science teaching, vol. 27, pp. 461- 476, 1990.

[5] Thomas R Koballa, "Attitude and relatedconcepts in science education," Science Education, vol. 72, no. 2, pp. 115-126, 1988.

[6] Abdeljalil Métioui and Louis Trudel. (2009) Revue permanente en ligne des utilisateurs des technologies de l'information et de la communication. [Online]. HYPERLINK "http://isdm.univ-tln.fr/PDF/isdm39/Article_Isdm_Ticemed09_Metioui_Trudel_PhB%20ok.pdf" http://isdm.univ-tln.fr/PDF/isdm39/Article_Isdm_Ticemed09_Metioui_Trudel_PhB%20ok.pdf

[7] Fouad Chafiqi and A Alagui , "Réforme éducative au Maroc et refonte des curricula dans les disciplines scientifiques," Carrefours de l'éducation, pp. 29-50, 2011.

[8] ABD-EL-KHALIC, F.; LEDERMAN, N.G, "Improving science teachers' conceptions of the nature of science: A critical review of the literature," International Journal of Science Education 22, pp. 665-701, 2000.

[9] Jacques Desautels and Marie Larochelle, Autour de l'idée de science.: De Boeck-Wesmael, 1992.

[10] ARIZA R PORLAN , GARCIA E GARCIA , and GARCIA RIVERO , "Porlan Ariza, P., & al. (1998).Les obstacles à la formation professionnelle des professeurs en rapport avec leurs idées sur la science, l'enseignement et l'apprentissage," Aster, pp. 207-235, 1998.

[11] Ezio Roletto, "La science et les connaissances scientifiques," Aste, no. 26, pp. 11-30, 1998.

[12] Rajaa AOULAD MOUSSA, Mourad MADRANE, Rachid JANNATI-IDRISSI, and Mouhamed LAAFOU, ENJEUX DE LA FORMATION INITIALE DES FUTURS ENSEIGNANTS DES SCIENCES AU PRIMAIRE, Avril 8-9, 2016."unpublished"

[13] Duit Reinders and David F Treagust, "Conceptual change: A powerful framework for improving science teaching and learning," International Journal of Science Education, vol. 25, no. 6, pp. 671-688, 2003.

[14] Commission Européenne EUR22845, "- L'enseignement scientifique aujourd'hui: une pédagogie renouvelée pour l'avenir de l'Europe Luxembourg," Office des publications officielles des Communautés européennes, 2007.

[15] D. HODSON, "In search of a meaninful relationship: an exploration of some issues relating to integration in science and science education," International Journal of Science education, pp. 541-562, 1992.

[16] Melody M Tsang, Ho Shu-Chun, and Liang Ting-Peng, "Consumer Attitudes Toward Mobile Advertising:An Empirical Study," International Journal of Electronic Commerce, vol. 8, no. 3, pp. 65–78, 2004.

[17] Mohamed Chekour , Mohaamed Al Achhab, and Mohamed Laafou, "Integration of Blended Learning in Teaching Computer Science in Moroccan High Schools," International Journal of Computer Technology and, vol. 4, no. 6, 2013.

[18] Marc Walckiers and Thomas De Praetere, "L'apprentissage collaboratif en ligne, huit avantages qui en font un must. Distances et savoirs," vol. 2, no. 1, pp. 53-75, 2004.

[19] J BENNETT, "Science with attitude: the perennial problem of pupils' responses to science," School Science Review, pp. 59-70, 2001.

[20] Jacques Desautels and Marie Larochelle, "Qu'est-ce que le savoir scientifique?," de la Pensée Éducative, pp. 96-98, 1993.

[21] "Relationships between a teacher's background and sensed adequacy to teach elementary science," Journal of Research in Science Teaching, vol. 12, pp. 85-88, 1975.

[22] Sébastien George, Gaëlle Molinari, Chihab Cherkaou, "Le réseau social Face-book comme support d'apprentissage pour les étudiants universitaires.," in 7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain, Agadir, 2015, pp. 102-113.

[23] LAROCHELLE M.in collaboration with PÉPIN Y DÉSAUTELS J., "Étude de la pertinence et de la viabilité d'une stratégie de formation à l'enseignement des sciences. ," Ottawa, 1994.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

A Real Time Embedded System Architecture for Autonomous Underwater Sensors Localization

[1]Redouane Es-sadaoui, [2]Jamal Khallaayoune, [2]Tamara Brizard
[1]National Institute of Posts and Telecommunications, Rabat, Morocco
[2]Aspremont, France
red.essadaoui@gmail.com, laayoune@inpt.ac.ma, tamara.brizard@arkeocean.com

ABSTRACT

Underwater Acoustic Sensor Networks (UWASNs) consist of a variable number of autonomous sensors or vehicles that are deployed over a given area to perform smart sensing and collaborative monitoring tasks. In UWASNs, sensor localization plays a critical role. Motivated by the advent of embedded systems and their widespread adoption in localization, this paper presents the design and architecture of an autonomous embedded system, that uses acoustic signal to communicate underwater. The proposed architecture implements a set of embedded interfaces, such as inter-processor communication link and serial interfaces, which facilitates its integration with other systems. The implementation of a straightforward localization algorithms based on the Phase Difference and the Time of Arrival techniques is also described. The ability of the developed system to localize underwater sensors was tested during sea trials.

Keywords- Underwater Acoustic Sensor Networks; localization; smart sensors; embedded system

1 Introduction

The majority of the earth's surface is covered by sea. The emergence of UWASNs provides new opportunities to explore the ocean in a variety of civilian and military applications thanks to their relative ease of deployment since they eliminate the need for cables, and reduce interfere with shipping activities. UWASNs will find applications in oceanographic data collection, pollution monitoring, oil and gas exploration and disaster prevention, and many more.

Underwater autonomous vehicles (AUVs), equipped with sensors, are a key enabling technology for the exploration of natural undersea resources and gathering of scientific data in collaborative monitoring missions. However, sensed data can only be integrated meaningfully when referenced to the location of the sensors, giving localization an important role. In addition, this may be also critical for sensor security, monitoring and recovery operations.

Various localization algorithms have been studied in terrestrial sensor networks, however, they cannot be applied directly to UWASNs because of attenuation of Global Positioning System (GPS) and Radio Frequency (RF) signals underwater, which makes the localization of UWASNs particularly challenging [1]. Thus, acoustic signals are used in underwater communications instead of GPS and RF signals, because they are less attenuated and travel further in water.

DOI: 10.14738/tmlai.54.3224
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3224

The advent of embedded systems is considered one of the enabling technologies for underwater localization. This paper proposes an experimented embedded system that uses pure tone narrowband acoustic signals to communicate between sensors immersed underwater. By measuring the phase difference of incoming pulses, the direction to the sensor can be calculated. The distance to the sensor is calculated based on the Time of Arrival of pulses.

Experiments have been conducted at sea to evaluate the ability of our system to localize sensors in an aquatic environment. Improvement techniques are envisaged to increase both distance and direction accuracy by optimizing the hardware architecture, software implementation, and integration of high performance localization algorithms.

The rest of paper is organized as follows. Section II presents localization methods from the literature. In Section III, the design and architecture of the embedded system are described. In Section IV, we explain the principle of adopted algorithms for sensor direction and distance estimation, followed by the experiments results at sea (Section V). Finally, in Section VI, a conclusion is presented.

2 Related works

Recently, many researches carried out for underwater acoustic sensor localization in UWASNs. For instance, in the paper titled "AoA Assisted Localization for Underwater Ad-Hoc Sensor Networks"[2], H. Huang et al proposed an angle of arrival (AoA) assisted localization scheme for underwater Ad-Hoc sensor networks in 2-D and 3-D. They showed the ability of this technique to improve accuracy in high localization coverage, for high performed distance estimation between sensor nodes and anchor nodes, in 2D and 3D sensor networks.

K. H. Lee et al [3] presented a study of underwater sensor's localization in order to extend this domain by developing a new algorithm based on terrestrial smart sensors network. The authors discussed some well-known methods for localization and distance calculation : Time of Arrival (ToA) that estimates the distance between nodes by calculating the propagation time of a signal, and trilateration technique that calculates the sensor position by intersection of three circles. Then, they proposed a method for localization by considering three initial unknown nodes randomly. The algorithm performance was examined by simulation using different conditions.

The paper [4], by V. M. Bhoopathy et al, presented a distributed wireless sensor network localization estimation based on two-way ToA method in order to avoid the need of time synchronization. This method is divided into two steps : The first step is anchor sensors localization (based on buoys position that are already fixed and their position is simple to know). The second step is sharing the anchors positions to the ordinary sensors, and running localization and mobility algorithm to ensure that the position is correct by comparing the detected position with the previous one and to reduce the effect of water current.

S. Kim and Y. Yoo [5] studied the characteristics of water movement and possible errors in Time Synchronization techniques. The authors defined a synchronization and localization schemes that improve accuracy, using Kalman and averaging filters. An inertial navigation system was adopted to eliminate localization errors caused by water current and propagation problems. This technique was evaluated by simulations and showed a reduced time synchronization latencies.

In [6] Z. Mousavi et al. presented a new method combining between localization and synchronization for communication in Underwater Sensors Networks. In this algorithm, the attribute of sound waves was considered and the stratification effect was used to compensate the bias in rang estimates. This method evaluates the precision of localization based on multiple factors: depth and temperature and salinity. Simulation results showed the effectiveness of the proposed method.

Typically, papers for sensor localization in UWASNs focus on two isolated parts of the localization process. The first part is the methods for estimating sensor ranges and direction, as in [7] [8], and the second part is the mathematical techniques and algorithms for determining the position of sensors in an entire UWASNs [9] [10].

3 System Architecture

3.1 Hardware architecture

Three architectures have been designed: 1) Acoustic Transmitter, integrable to sensors only transmitting, 2) Acoustic receiver to localize transmitters 3) Combined acoustic transceiver, able to work in both transmitter or receiver modes. This multi-choise architecture provides a low cost, low power consumption and an ease-to-use solution, that can be integrated in both pre-existing or new sensor networks infrastructures, depending on the application requirements.

3.1.1 Acoustic transmitter

The acoustic transmitter, presented in Fig. 1, is based on an ultra-low-power consumption flash microcontroller (Flash MCU) that features a powerful 16-bit Reduced Instruction Set Computing "RISC" architecture, 16-bit registers, and constant generators that contribute to maximum code efficiency.

The Flash MCU outputs acoustic pulses in 8-bits digital format, the pulse is converted to an analog format by a Digital to Analog Converter (DAC), then amplified to high amplitude (around 24 volts) by means of an analog converter. A transformer is installed after the amplifier to raise the voltage of the acoustic pulse to a high amplitude voltage, which will attack an acoustic transducer ceramic that convert the electrical wave to an acoustic wave propagating in the water.

The Flash MCU also manages communications on the surface through an RF transceiver. A set of communication interfaces like Universal Serial Bus (USB), Serial Peripheral Interface (SPI), Inter-Integrated Circuit (I2C), Universal Asynchronous Receiver/Transmitter (UART) and Ethernet are provided on the Flash MCU to communicate with external hosts and sensors.

3.1.2 Acoustic receiver

The receiver electronics, presented in Fig. 2, incorporates a fixed-point digital signal processor (DSP) based on an advanced modified Harvard architecture which provides an arithmetic logic unit (ALU) with a high degree of parallelism, application-specific hardware logic, on-chip memory, and additional on-chip peripherals [11].

A pre-amplifier and low pass filter have been designed to answer to the specified frequency band and signal amplitude, thus, the acoustic pulses detected on the hydrophones array antenna, which is formed from at least two hydrophones, are first amplified, filtered and converted to digital format through stereo analog-to-digital converters (ADCs), that perform sampling, analog to-digital conversion, and anti-alias filtering.

The digital output of ADCs are communicated to DSP through a Multichannel Buffered Serial Ports (McBSP) [12]. The clock generator supplies the DSP and ADCs with the required clocks. The Flash MCU manages the RF transceiver, the communication to the DSP using Host Port Interface (HPI), and communication to an external host and sensors .

3.1.3 Acoustic tranceiver

This architecture, presented in Fig. 3, combines both transmitter and receiver on the same board, giving the system the ability to play as transmitter or receiver as needed (Fig .3).

3.2 Enabling features

3.2.1 Acoustic pulse form

The transmitted pulse is a pure tone sine wave, shaped in a blackman window. The trasnmitted pulse is obatined by multipling the sine wave by a blakman window.

Equations (1) and (2) define respectively the equations of the blackman window and the transmitted pulse.

Where A is the signal amplitude peak, f is the signal frequency, fe is the sampling frequency and N is the length of the blackman window.

$$\omega(k) = 0.42 - 0.5.\cos(2\pi k/(N-1)) + 0.08.\cos(4\pi k/(N-1)) \quad (1)$$

$$S(k) = A.\sin(2\pi.f/fe).\ \omega(k) \quad (2)$$

Fig. 4(a) shows the form of a transmitted pulse, simulated in matlab, for a signal frequency of 22KHz, a sampling frequency of 48KHz and a pulse length of 10 milliseconds.

The real output of the transmitter was measured and qualified by using an oscilloscope. Fig. 4(b) shows the blackman pulse measured at the output of the 24Volts amplifier.

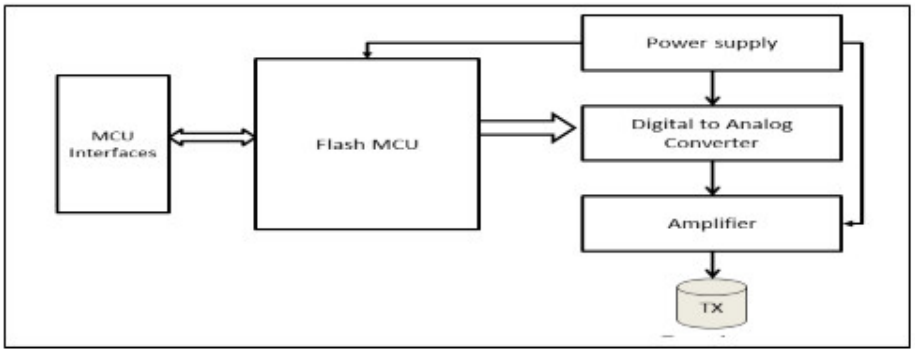


Figure 1. Acoustic transmitter architecture

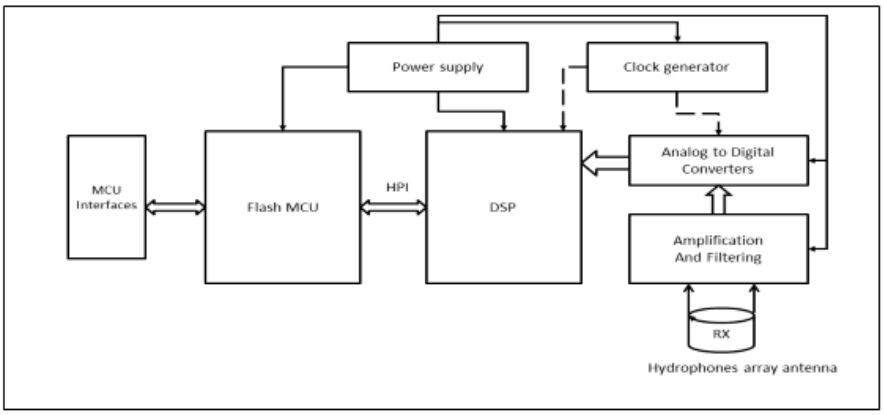


Figure 2. Acoustic receiver architecture

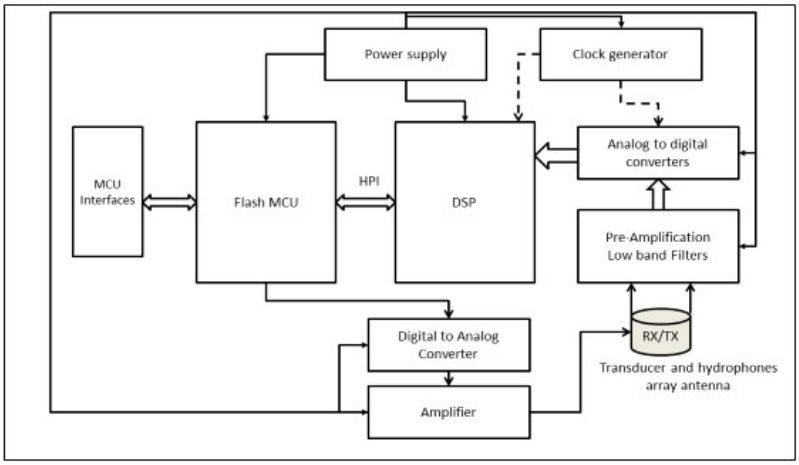


Figure 3. Acoustic transceiver architecture

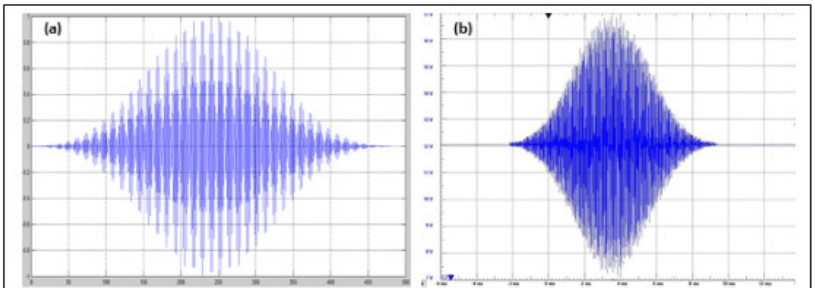


Figure 4. Simulated (a) and measured (b) acoustic pulses

3.2.2 Interprocessor communication through HPI lik:

The host port interface (HPI) is a parallel port through which the Flash MCU can directly access to DSP memory space. The Flash MCU device functions as a master to the interface, which increases ease of access [12].

The Flash MCU and DSP can exchange information via internal or external memory. A HPI software has been developed on the Flash MCU to perform inter-processor communication between the Flash MCU and DSP. The main C functions that were developed for HPI communications are:

- HPI_init(): Initialize HPI link
- HPI_read(): Read a data from DSP memory.
- HPI_write(): Write a data to DSP memory.
- HPI_isr(): Handle interrupts from DSP.
- HPI_dspInt(): Interrupt DSP through HPI link.

3.2.3 DSP Bootloader

The DSP bootloader is commonly used to transfer code from an external source into either internal or external memory. The bootloader allows the user flexibility in system design by storing code in cost-efficient, nonvolatile external memory. This enables the designer to avoid a custom on-chip ROM mask and to reduce system cost. It is possible for the Flash MCU to download code and bootload the DSP using the HPI.

In order to save memory, the DSP Firmware is stored in the Flash MCU memory. The Flash MCU utilizes the HPI software (described above) to perform a DSP bootloader using HPI link.

3.2.4 Serial and Ethernet links:

A Hardwired TCP/IP embedded Ethernet controller is interfaced to the Flash MCU. The software TCP/IP driver was developed on the Flash MCU to provide an easier TCP/IP connection to external hosts. A set of serial communication links are provided on the MCU external Interface to support standard communication protocols required by many digital sensors like GPS, compass, and accelerometer.

4 Sensor Localization

4.1 Direction estimation

A popular way to measure the direction (relative bearing) of a sound wave is by measuring the phase difference of the signals of hydrophones array. Fig. 5 presents the block diagram of the phase difference chain, implemented in DSP.

Fig. 6 shows the simulation of relative bearing computing in Matlab, based on the phase difference of two sine pure tone acoustic signal, of frequency 20KHz, and Signal Noise Ratio of 12decibels. The Sampling frequency is taken equal to 48 KHz.

The simulation results shows that the accuracy deteriorates slightly for a phase difference close to 90 degrees, and becomes best as the phase difference goes far from 90 degrees.

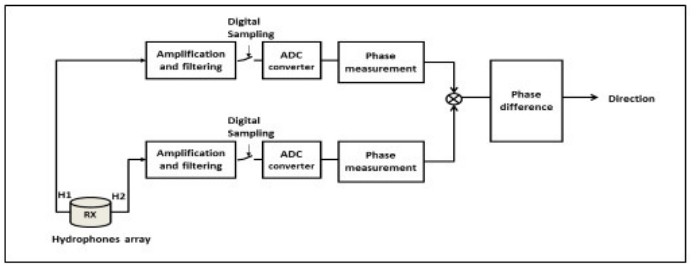


Figure 1. Direction measurement block diagram

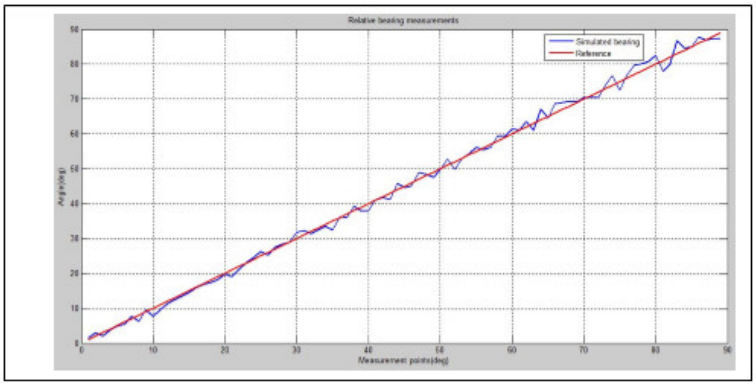


Figure 6. Direction computing simulation, for a relative bearing from 0 to 90 degrees

4.2 Distance estimation

We assume that the transmitter and the receiver are time synchronized. Knowing the sound velocity (Svelocity) of water, the distance (d) to the transmitter is calculated by the receiver by listening for the pulses sent by the transmitter and calculating the Time of Arrival (ToA) to estimate the propoagation time of detected pulses (3). Fig. 7 shows the block diagram of the distance computing, implemented in DSP.

$$d = (TOA * S_{velocity}) \quad (3)$$

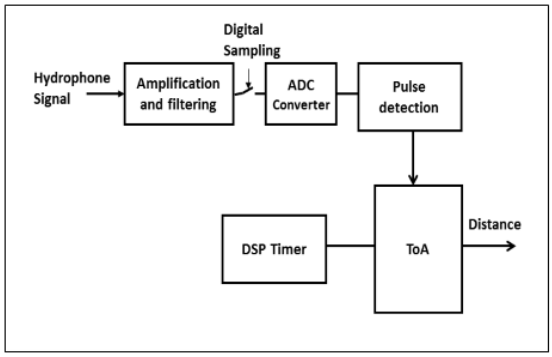


Figure 7. Distance measurement block diagram

4.3 Position estimation

We assume that the angle to the north is known. The position of a sensor integrating the acoustic transmitter device or the acoustic transceiver device can be determined by an acoustic receiver with known coordinates like surface vessel and anchor sensors. Sensors integrating the acoustic receiver device or the acoustic transceiver device can estimate their positions in reference to transmitters installed in known locations.

Figure below depicts how the position of a sensor in 2D space (Xp, Yp) is determined in reference to a fixed point (sensor 1) with known coordinates (x1, y1). d is the measured distance between the sensor and the reference point, α is the measured direction (4)(5).

$$Xp = x1 - d * \cos(\alpha) \quad (4)$$

$$Yp = y1 - d * \sin(\alpha) \quad (5)$$

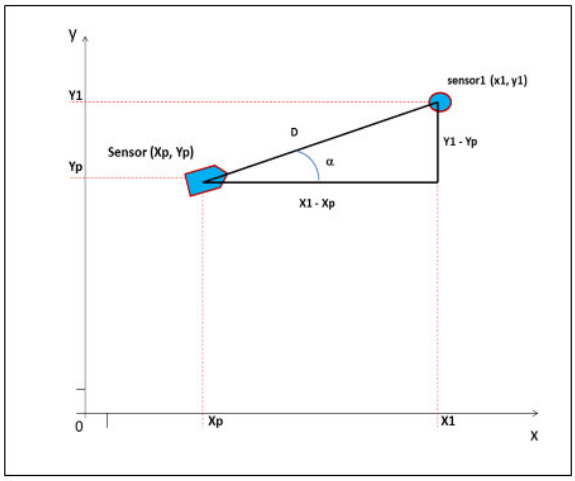


Figure 8. Location estimation using distance and relative bearing

5 Sea Experiments

5.1 Experimental setup

Extensive sea trials were performed in BOUREGREG MARINA, located at the mouth of the Bouregreg River, on the shore of SALE, Morocco. This location provides an easy-to-access environment for sea trials (Fig. 9).

The transmitter electronic board was set on the deck, while its watertight ceramic was immersed at 2 meters depth. The receiver electronic board, connected to a Laptop, was also set on the deck, 40 meters far away from the transmitter. The receiver phased array antenna was immersed at a depth 2 meters and fixed to a 360 degrees turntable, which allows the user to evaluate the relative bearing accuracy of the receiver in comparison to the reference angle set through the turntable. The transmitter.

Figure 9. Sea water test environment and setup

5.2 Experimental results

The experimental results of relative bearing accuracy, have been performed with an acoustic frequency of 22KHz. These results have been obtained by varying the reference angle with a minimum step of 6 degrees, 15 records per step, and logging the angle (relative bearing) calculated by the receiver.

Fig. 10 presents the obtained measurement results, for a relative bearing between 360 degrees and 260 degrees.

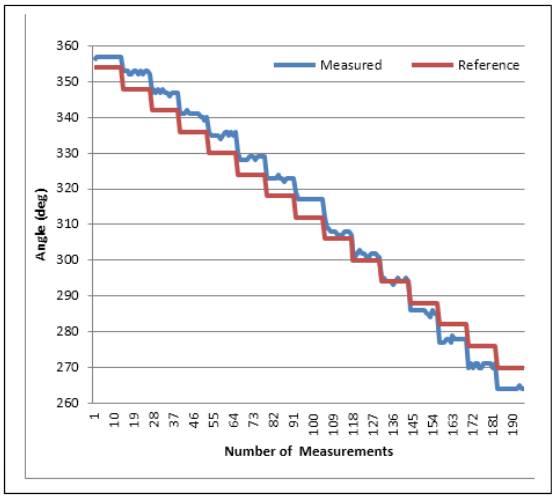


Figure 10. Sea water direction measurements

The sea experiments result shows that the receiver calculates correctly the direction to the transmitter with an estimated error of less than +/-10 degrees. Distance measurements have been performed within a maximum range of 400 meters, they showed an average distance error of around +/-2 meters.

6 Conclusion

We have designed, developed and qualified an autonomous real time embedded system, able to localize acoustic sensors in an underwater environment. The developed system can be considered as an enabling technology for practical experimentations of underwater localization approaches that are currently limited to laboratory and simulation.

In this paper, we have presented the functional design and implementation of a straightforward localization method, based on the phase difference scheme to estimate the direction and the Time of Arrival for distance computing.

The real performances were verified by sea trials, showing less than +/-10 degrees error in the direction, and around +/-2 meters distance error. The obtained results are considered as satisfactory for many UWASNs applications, however, more accurate localization system is the subject of our future research.

REFERENCES

[1] Akyildiz, I.F.; Pompili, D.; Melodia, T. Underwater acoustic sensor networks: Research challenges. Ad Hoc Netw. 2005, 3, 257-279 .

[2] Kang Hoon Lee, Chang Ho Yu, Jae Weon Choi and Young Bong Seo, "ToA based sensor localization in underwater wireless sensor networks," 2008 SICE Annual Conference, Tokyo, 2008, pp. 1357-1361. doi: 10.1109/SICE.2008.4654869

[3] Huai Huang and Yahong Rosa Zheng, "AoA assisted localization for underwater Ad-Hoc sensor networks," OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, 2016, pp. 1-6. doi: 10.1109/ OCEANS.2016.7761388

[4] V. Mandalapa Bhoopathy, M. Ben Haj Frej, S. Richard Ebenezer Amalorpavaraj and I. Shaik, "Localization and mobility of underwater acoustic sensor nodes," 2016 Annual Connecticut Conference on Industrial Electronics, Technology & Automation (CT-IETA), Bridgeport, CT, USA, 2016, pp. 1-5. doi: 10.1109/CT-IETA.2016.7868249

[5] Sungryul Kim, Younghwan Yoo, "SLSMP: Time Synchronization and Localization Using Seawater Movement Pattern in Underwater Wireless Networks". nternational Journal of Distributed Sensor Networks. Vol 10, Issue 1, 2014 10.1155/2014/172043.

[6] MOUSAVI, ZAHRA and REZA JAVIDAN. "A Combined Localization-Synchronization Method for Underwater Communication." International Journal of Computer Networks and Communications Security. VOL. 2, NO. 12, DECEMBER 2014, 414–422. ISSN 2308-9830

[7] Y. Xu, W. Dandan and F. Hua, "Underwater acoustic source localization method based on TDOA with particle filtering," The 26th Chinese Control and Decision Conference (2014 CCDC), Changsha, 2014, pp. 4634-4637.

[8] P. Jiang, "An underwater sensor network localization algorithm based on Time-of-Arrivals (TOAs)," 2013 Ninth International Conference on Natural Computation (ICNC), Shenyang, 2013, pp. 1552-1556. doi: 10.1109/ICNC.2013.6818228

[9] S. Zhu, N. Jin, L. Wang, X. Zheng, S. Yang and M. Zhu, "A novel dual-hydrophone localization method in underwater sensor networks," 2016 IEEE/OES China Ocean Acoustics (COA), Harbin, 2016, pp. 1-4. doi: 10.1109/COA.2016.7535787

[10] C. H. Yu, K. H. Lee, Hyun Pil Moon, J. W. Choi and Y. B. Seo, "Sensor localization algorithms in underwater wireless sensor networks," 2009 ICCAS-SICE, Fukuoka, 2009, pp. 1760-1764.

[11] Texas Instruments, "TMS320C54x DSP Functional Overview". September 1998 – Revised May 2000.

[12] Texas Instruments, "TMS320C54x DSP Reference Set, Volume 5: Enhanced Peripherals". March 2007.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Computing on Encrypted Data into the Cloud though Fully Homomorphic Encryption

Samiha Jlilab, Hassan Satori, Khalid Satori
LIIAN, Department of Mathematics and Computer Science, FSDM, USMBA,
Fez, Morocco
Department of Mathematics and Computer Science, FPN, UMP
Nador, Morocco
samihajlilab00@gmail.com, hsatori@yahoo.com

ABSTRACT

Securing Data in the cloud based on Fully Homomorphic Encryption (FHE) is a new and potential form of security that allows computing on encrypted data without decrypted it first. However, a practical FHE solution is not available for implementation today. In this work, we propose a platform based on open source solutions to perform data computations (addition and multiplication) on encrypted form. In addition, taking account of efficiently and the security component, the most popular partially homomorphic encryption algorithms (RSA, Paillier and ElGamal) are studied to analyze the process times of encryption, decryption and computation of each algorithm. Furthermore, to compromise between performance and security, we need to study different key sizes and different data sizes as parameters.

Keywords: Cloud Computing, Third-Party;Data Privacy, Fully Homomorphic Encryption, Partially Homomorphic Encryption, RSA, ElGamal, Paillier, Gentry's scheme.

1 Introduction

Cloud computing holds enormous potentials (cost reduction, shared computing-resources, elasticity, good performance, etc.). As a matter of fact, the cloud has few challenges like any other innovative technology. Data security is a key challenge when moving to the cloud[1]. In fact, when you move your data to the cloud, you are losing control. Cloud gives access to data, but you have no way to make sure that someone else does not have access to it as well. In addition, exposing your data in a public cloud environment shared with other companies may put your data at risk and may you lose the confidentiality of your sensitive data.The key solution is tostore and perform computations on cipher-text over a public area. This goal can be achieved using the homomorphic encryption proprieties[2]. Homomorphic encryption is the mathematical process of performing calculations on encrypted data which cannot be read or analyzed by anyone other than the client who retains the encryption and decryption keys. This kind of cryptography is an efficient solution to protect from internal and externals threats (curious cloud administrator, malicious cloud users, attackers).

Recently, many researchers are interested in fully homomorphic encryption solutions to improve data privacy and eliminate trusted third parties i.e. performing computational operations on encrypted data without learning anything about it by a third party. Naveed Islam et al[3] have exploited the multiplicative

DOI: 10.14738/tmlai.54.3225
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3225

and the additive homomorphic proprieties of RSA and Paillier algorithms to secure image sharing between players. M.Tebaa et al. [4-5] have implemented the two previous asymmetric algorithms, on a VMware vSphere environment, to secure computations on sensitive data (bank data). A.Chantterjee and I.Sengupta [6]applied FHE technique to search and store encyrpted data on the cloud environnement and they have re-used the homomorphic modules proposed in scarab library [7]to perform search and store over encrypted data.

Our contribution is to develop a practical, efficient and secure platform based on different homomorphic algorithms (Paillier, RSA and ElGamal) to allow performing calculations, by a third party, with encrypted data. The performance of our platform is analyzed by studying the processing time of encryption, computation and decryption phases.

The rest of this paper is divided into seven sections. In the 2nd section, we start by introducing cloud computing concept and the security component as issue for cloud adaptation by companies and users. In the 3rd section, we define the homomorphic encryption propriety, its types and some examples of homomorphic cryptosystems. In the 4th section, we present the system conception. The experimental platform is described in the 5th section. In the 6th section we present our simulation procedure and results. Finally, the conclusion and perspectives will be mentioned in 7th section.

2 Cloud Computing

Cloud Computing is the new form of Information Technology (IT) outsourcing that provides its solutions as an on-demand consumable services. It consists of using applications, platforms or distributed virtual infrastructures which are not necessarily located in the company's premises. The value of Cloud computing is to introduce a new way of managing IT, providing better control of Direction of information systems (DIS) expenses and allowing enterprises to renovate the core of their business without worrying about time and money constraints related to the integration of new technologies.

2.1 Official Definition of Cloud Computing

The National Institute of Standards and Technology (NIST)defined cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction[8].

The NIST definition lists five essential characteristics of Cloud Computing: on-demand self-service, broad network access, resource pooling, rapid elasticity or expansion, and measured service. It also lists three service models (IaaS, PaaS and SaaS), and four deployment models (public, private, community and hybrid).

2.2 Security issue in cloud computing

The NIST's cloud definition did not include security such as an essential characteristic of the cloud computing however security is a potential concept which must intervene for preserving and protecting the confidential and sensitive data of each enterprises. In fact, every technology has a flaw so does cloud computing. The client has also participate to protect his data from harm (attackers, malicious administrator…). So, cloud security should be a shared responsibility between cloud provider and cloud users.

Cloud security requires a profound questioning of corporate security policies. They must go beyond the narrow managing passwords and login privileges (access control). It is necessary to take the next step and think safety in terms of use and data types [9]. For this, Fully Homomorphic Encryption will be a good solution to protect the data and an efficient way to take advantage of cloud services.

3 Homomorphic Encryption

In traditional encryption schemes, Bob encrypts a plaintext message to obtain a cipher text. Alice decrypts the cipher-text to find the plaintext. So, with this standard encryption, we can only secure the following steps: establishment of communication and data transfer. In Homomorphic Encryption, parties that do not know anything about the plaintext data can perform computations on it by performing computations on the corresponding cipher-text.

Definition of Homomorphic Encryption:

The encryption algorithm E () is homomorphic if given E(x) and E(y), one can obtain E (x□y) without decrypting x; y for some operation □ (+, ×...) [10].

3.1 Partially Homomorphic Encryption (PHE):

A cryptosystem is partially homomorphic if it supports adding or multiplying of cipher-texts but not both operations at the same time. The Goldwasser-Micali [11] and Paillier [12] schemes supported addition operations, while the RSA [13] and ElGamal [14]schemes supported multiplication operations.

3.1.1 Additive Homomorphic Encryption (AHE):

A scheme is an additive Homomorphic Encryption (AHE) if you give only the public-key and the encryption of $\mathrm{m1}$ and$\mathrm{m2}$, one can compute the encryption of$\mathrm{m1} + \mathrm{m2}$.

Mathematically, E is an AHF if:

$$E(x + y) = E(x) \times E(y)$$

$$E\left(\sum_{i=1}^{n} m_i\right) = \prod_{i=1}^{n} E(m_i)$$

3.1.2 Multiplicative Homomorphic Encryption (MHE):

A scheme is a Multiplicative Homomorphic Encryption (MHE) if you give only the public-key and the encryption of $\mathrm{m1}$ and$\mathrm{m2}$, one can compute the encryption of$\mathrm{m}_1\mathrm{m}_2$.

Mathematically, E is a MHE if:

$$E(x \times y) = E(x) \times E(y)$$

$$E\left(\prod_{i=1}^{n} m_i\right) = \prod_{i=1}^{n} E(m_i)$$

3.2 Asymmetric Homomorphic Encryption Algorithms

3.2.1 Paillier Scheme; an AHE:

Paillier cryptosystem is an asymmetric algorithm for public key cryptography and it is an AHE.The algorithm consists of three components: the key generator, the encryption algorithm and the decryption algorithm. The three components work as follow:

Table 1. Paillier Cryptosystem

Key Generator : KeyGen(p, q)
Input: $\boldsymbol{p, q} \in \mathbb{Z}, \boldsymbol{\gcd(p,q) = 1}$
Compute $\boldsymbol{n = p \times q}$, $\boldsymbol{\lambda = lcm(p-1, q-1) = \frac{(p-1)(q-1)}{pgc(p-1,q-1)}}$ Choose $\boldsymbol{g} \in \mathbb{Z}^*_{n^2}$ such that $$\boldsymbol{\gcd\left(L\left(g^{\lambda} mod n^2\right), n\right) = 1 with L(u) = \frac{u-1}{n}}$$
Output: $(\boldsymbol{pk}, \boldsymbol{sk})$ Public key: $\boldsymbol{pk} = (\boldsymbol{n}, \boldsymbol{g})$ Secret key: $\boldsymbol{sk} = (\boldsymbol{p}, \boldsymbol{q})$
Encryption: E(m, pk)
Input: $\boldsymbol{m} \in \mathbb{Z}_{\boldsymbol{n}} with \boldsymbol{m < n}$
Choose $\boldsymbol{r} \in \mathbb{Z}^*_{n^2}$ Compute $\boldsymbol{C = g^m r^r mod(n^2)}$
Output: $\boldsymbol{C} \in \mathbb{Z}_{n^2}$
Decryption: D(c, sk)
Input: $\boldsymbol{C} \in \mathbb{Z}_{n^2}$
Compute $\boldsymbol{m = \frac{L(c^{\lambda} mod n^2)}{L(g^{\lambda} mod n^2)} mod n}$
Output: $\boldsymbol{m} \in \mathbb{Z}_{\boldsymbol{n}}$

Paillier's algorithm realizes the property of additive homomorphic encryption.

Proof: suppose we have two ciphers $E(m_1) = C_1 et E(m_2) = C_2$

$$E(m_1) = C_1 = g^{m1}.r_1^n mod n^2$$

$$E(m_2) = C_2 = g^{m2}.r_2^n \ mod \ n^2$$

So:

$$C_1.C_2 = g^{m1}.r_1^n \times g^{m2}.r_2^n \ mod \ n^2 = g^{m1+m2}.(r_1 r_2)^n \ mod$$

i.e.

$$\boldsymbol{E(m_1).E(m_2) = E(m_1 + m_2)}$$

3.2.1 Paillier cryptosystem can be a MHE?

Suppose we have two plaintexts m_1 et m_2:

$E(m_1) = g^{m1}.r_1^n mod n^2$.

$$E(m_1)^{m_2} = \ g^{m1\,m_2}.r_1^n \ mod \ n^2 = E(m_1 m_2).$$

So with this trick, we can realize multiplicative operation with Paillier cryptosystem.

3.2.2 RSA Scheme, a MHE:

RSA is an asymmetric cryptosystem. It was described in 1977 by Ron Rivest, Adi Shamir and Leonard Adleman. The scheme summarizes in three steps: the key generator, the encryption algorithm and the decryption algorithm.The three components work as follow:

Table 2 RSA cryptosystem

Key Generations : KeyGen(p, q)
Input: $\boldsymbol{p, q} \in \mathbb{Z}, \boldsymbol{\gcd(p, q) = 1}$
Compute $\boldsymbol{n = p \times q}$ $\boldsymbol{\varphi(n) = (p-1)(q-1)}$ Choose $\boldsymbol{e} \in \mathbb{Z}_{\boldsymbol{n}}$ such that $\boldsymbol{\gcd(e, \varphi(n)) = 1}$ Determine $\boldsymbol{d} \in \mathbb{Z}_{\boldsymbol{n}}$ such that $\boldsymbol{e \times d = 1\ mod(\varphi(n))}$
Output: $(\boldsymbol{pk, sk})$ Public key: $\boldsymbol{pk = (n, e)}$ Secret key: $\boldsymbol{sk = d}$
Encryption: E(m,pk)
Input: $\boldsymbol{m} \in \mathbb{Z}_{\boldsymbol{n}} \boldsymbol{with m < n}$
Compute $\boldsymbol{C = m^e mod(n)}$
Output: $\boldsymbol{C} \in \mathbb{Z}_{\boldsymbol{n}}$
Decryption: D(c. sk)
Input: $\boldsymbol{C} \in \mathbb{Z}_{\boldsymbol{n}}$
Compute $\boldsymbol{m = C^d mod n}$
Output: $\boldsymbol{m} \in \mathbb{Z}_{\boldsymbol{n}}$

RSA realizes the property of Multiplicative Homomorphic encryption.

Proof: Suppose we have two hers $E(m_1) = C_1 et E(m_2) = C_2$

$$E(m_1) = C_1 = {m_1}^e mod(n)$$

$$E(m_2) = C_2 = {m_2}^e mod(n)$$

$$E(m_1) \times E(m_2) = {m_1}^e {m_2}^e mod(n) = (m_1 m_2)^e mod(n)$$

And $E(m_1 m_2) = (m_1 m_2)^e mod(n)$

So: $\boldsymbol{E(m_1) \times E(m_2) = E(m_1 m_2)}$

3.2.3 ElGamal cryptosystem; a MHE:

The ElGamal encryption system is an asymmetric key encryption algorithm for public-key cryptography which is based on the Diffie-Hellman key exchange. It was described by Taher El-Gamal in 1985.

The scheme summarizes in three steps which are described in table3.

Table 3 ElGamal cryptosystem

Key Generator : KeyGen(p, g, s)
Input: $\boldsymbol{p} \in \mathbb{Z}, \boldsymbol{g} \in \mathbb{Z}_p$
Choose $\boldsymbol{s} \in \mathbb{Z}, \boldsymbol{with}\ \boldsymbol{1 < s < p - 1}$ Choose $\boldsymbol{h} \in \mathbb{Z}_n$ Compute $\boldsymbol{h = g^s}$
Output: $(\boldsymbol{pk}, \boldsymbol{sk})$ Public key: $\boldsymbol{pk} = (\boldsymbol{p}, \boldsymbol{g}, \boldsymbol{h})$ Secret key: $\boldsymbol{sk} = \boldsymbol{s}$
Encryption: E(m, pk)
Input: $\boldsymbol{m} \in \mathbb{Z}_p$
Compute $\boldsymbol{E(m) = (C1, C2) = (g^k, mh^k)\ mod(p)}$
Output: $\boldsymbol{C1},\ \boldsymbol{C2} \in \mathbb{Z}_p$
Decryption: D(c, sk)
Input: $\boldsymbol{C1},\ \boldsymbol{C2} \in \mathbb{Z}_p$
Compute $\boldsymbol{m = \frac{C2}{C1^s} modpi.e.\ m = C_1{}^{p-1-s}.C_2 modp}$
Output: $\boldsymbol{m} \in \mathbb{Z}_p$

The scheme realizes the property of MHE.

Proof: Suppose we have two ciphers $E(m_1) et E(m_2)$

$$E(m_1) = (g^{k1}, m_1 h^{k1})$$

$$E(m_2) = (g^{k2}, m_2 h^{k2})$$

So: $E(m_1) \times E(m_2) = \left(g^{k1}, m_1 h^{k1}\right) \times \left(g^{k2}, m_2 h^{k2}\right)$

$= (g^{k1+k2}, m_1 m_2 \times h^{k1+k2})$.

And: $E(m_1 m_2) = (g^k, m_1 m_2 \times h^k)$.

⇨ $E(m_1) \times E(m_2) = E(m_1 m_2)$.

3.2.4 ElGamal cryptosystem can be an AHE?

If we practice some modifications on El-Gamal scheme such as replacing the plaintextmbyg$^{\mathrm{m}}$, this scheme will be realized the property of an additive homomorphic cryptosystem.

Proof: Suppose we have two ciphers$E(m_1) et E(m_2)$:

$$E(m_1) = (g^{k1}, g^{m_1}.h^{k1})$$

$$E(m_2) = (g^{k2}, g^{m_2}.h^{k2})$$

So:

$$E(m_1) \times E(m_2) = \left(g^{k1}, g^{m_1} h^{k1}\right) \times \left(g^{k2}, g^{m_2} h^{k2}\right)$$

$$= (g^{k1+k2}, g^{m_2+m_1} \times h^{k1+k2})$$

⇨ $E(m_1) \times E(m_2) = E(m_1 + m_2)$

3.3 Fully Homomorphic Encryption(FHE)

Fully Homomorphic Encryption (FHE) allows performed arbitrary operations on encrypted data without actually observing the raw data.

Being fully homomorphic means that whenever f is a function composed of additions and multiplications in the ring, then $\mathrm{Dec}(\mathrm{f}(\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n}) = \mathrm{f}(\mathrm{m}_1,\ldots,\mathrm{m}_\mathrm{n})$. So, if the cloud provider can efficiently compute $\mathrm{f}(\mathrm{m}_1,\ldots,\mathrm{m}_\mathrm{n})$ from cipher-texts $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n}$, without learning any information about the corresponding plaintexts$\mathrm{m}_1,\ldots,\mathrm{m}_\mathrm{n}$, then the system is efficient and secure [15].

In 2009, Craig Gentry has invented the first fully homomorphic encryption scheme (based on ideal lattices and LWE) that evaluates an arbitrary number of additions and multiplications and thus calculate any type of function on encrypted data [16].

Gentry's Algorithm contains the following four steps:

Key-generator: three keys $(\boldsymbol{S_k}, \boldsymbol{P_k}, \boldsymbol{E_k})$
Encryption: Enc (pk, m) → $\boldsymbol{C}$
Evaluation: $\boldsymbol{Eval_{E_k}(f, C)}$
Decryption: Dec($\boldsymbol{S_k}$,$\boldsymbol{C}$) → f(m)

Gentry's model includes several steps. The first one is constructing the Somewhat Homomorphic Encryption (SWHE) that supported a limited number of additions and multiplications on cipher-text in order to limit the noise component. The next step is squashing the decryption function to be expressed as a low degree polynomial. Then, using the Bootstrapping technique to refresh the cipher-texts before every homomorphic operation. Using this technique, the number of homomorphic operations becomes unlimited and we get a FHE scheme.

Gentry's solution improves the efficiency of secure multiparty computation. But, until today, his theoretical solution has a practical problem: it is complicated and imposes a large public keys size. Also, perform computations on encrypted data is too slow than on raw data[17].

3.3.1 DGHV Scheme;a FHE over integer

Based on Gentry's construction, V. Dijk et al[18] have proposed a simple algorithm applying simple addition and multiplication in order to reduce size of keys by using subset of the original public key. The table 5 as below, describe the various steps of DGVH scheme.

Table 4 DGVH scheme

Parameters
p: *secret key*, p *is a prime odd integer* $ev_k = \{x_i = pq_i + r_i\}$
Encryption
Input : $m \in \{0,1\}$ *compute* : $c = pq + 2r + m$
Evaluation
compute : $c_{add} = (c_1 + c_2)\ mod\ x_0$ *compute* : $c_{mul} = (c_1 c_2)\ mod\ x_0$
Decryption
Input: c *compute* : $c\ mod p\ mod 2 = (2r + m) mod 2 = m$ *Output*: m

3.3.2 BGV Scheme; FHE without bootstrapping

Z. Brakerski et al.[19] have proposed a new approach without gentry's bootstrapping technique. For managing the noise component of cipher-text as homomorphic operations are performed, the BGV is based on new techniques developed by Brakerski and vaikuntanathan[20]. The table 6 as below, describes the BGV scheme.

Table 5 BGV scheme

Parameters
λ: $secuity\ parameter$
n, k: $positives\ integers$
q : $odd\ integer$
e: $noise$
A: $k \times n\ marix \in Z_q^{k \times n}$
s: $secret\ key$
$v = As + 2e$: $public\ key$
Encryption
Input : $m \in \{0,1\}$, r $\in \{0,1\}^k$
Compute: $c_1 = A^T r and c_2 = v^T + m$
Output: $(c_1, c_2) \in Z_q^{n+1}$
Decryption
Input : (c_1, c_2)
Compute: $C = c_2 - < a, s >$
$m = C mod 2$
Output: $m \in Z_q$

All these algorithms can be applied on several domains and applications (e-voting System, Cloud Computing services, banking and Financial Applications, Medical Applications, Securing Treatment of Personal data, etc.)[21]–[24]

4 System conception

As we previously mentioned, this work focus on finding a scheme to treat encrypted data and perform complex calculations, in a cloud environment, without first decrypted it. So, the main goal is to create an algorithm on which we can use both cryptosystems when we have arbitrary operations. We can summarize our approach on three principal steps:

4.1 Encryption

we create, locally, a program based on PHE schemes to create an encrypted files or encrypted databases. In this program, we take into account the concept of parallelism because a parallel calculation is performed on more computing units. I.e. the computation is divided into parts that can run concurrently in order to reduce the processing time, solve problems of large size and to be able to handle multiple things at one time. Then, we send encrypted data into the cloud provider to store them.

4.2 Treatment

includes three sub-steps*:*

- Sends a request to cloud to perform calculations on encrypted data.
- Cloud's compute server has a function f for doing computations of cipher-texts. This computations are performed in order of priority.

- Send the result, of the request, which will be decrypted by Client Company.

4.3 Decryption:

Decrypts the result using the private key and comparing the results obtained with our results performed on raw data.

The Figure.1 presents the architecture of our conception system.

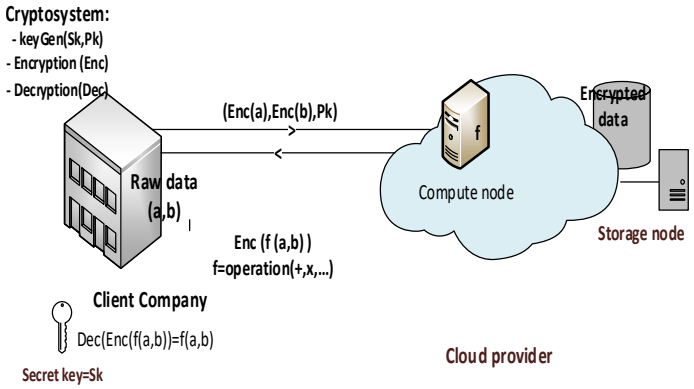


Figure 3. The proposed architecture of computing on encrypted data.

5 Experimental setup

The execution results are taken on Dell laptop having Intel® core ™ i5-4310M (2.70 GHz) Processor, 8 GB DDR3L (RAM), 320GB HDD, and Ubuntu 16.04 LTS Desktop (64-bits) operating system.

Concerning the platform of this work, our implementation is divided into two main parts:

5.1 Cloud platform:

using a cloud environment is a main part of our thesis. So, we implement a virtual cloud platform based on open-source solutions: OpenStack and KVM. The installation and the configuration of OpenStack's services implemented on four nodes (controller node, compute node, networking node and storage-Swift node). Each node is a VM under KVM having minimums material characteristics and the Ubuntu server 14.04 LTS 64 bits as operating system. We had tested our cloud functionality by creating network devices, creating images and instances, uploading and downloading raw data (files and images).

5.2 Java platform*:*

we use java language to test various homomorphic encryption algorithms in order to analyze and compare between them. This way, we're going to take account of the process times of encryption, computation and decryption time, size of keys and size of data. In addition, java platform provides an excellent base for writing secure applications by integrating JCA and JCE frameworks, and Bouncy Castle provider to implement the algorithms easily and with few lines of code. In this work, we implemented manually the cryptosystems in order to do some modifications on the encryption equations of Paillier and ElGamal algorithms.

The experiment is diagrammed as follow:

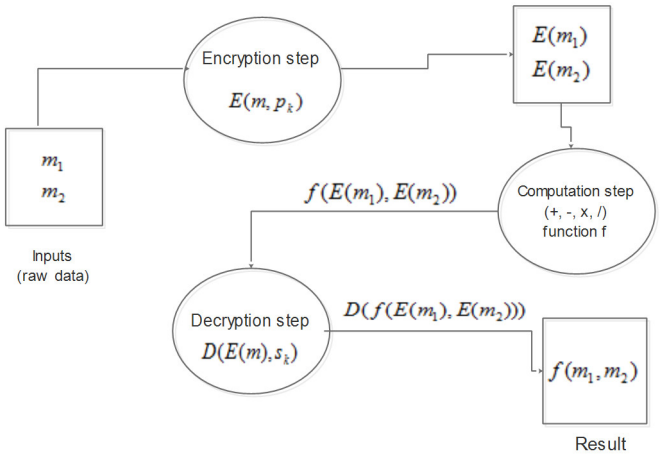


Figure 4. Diagram of homomorphic encryption operations processing.

6 Simulation procudure and results

6.1 Procedure

Our experimental starts by generating a set of data with different size (from 32bytes up to 1MB). In addition we are using digital data type to perform our mathematical calculations. In fact, the type of the data doesn't affect our results because encryption depends only on data size and not data type.

Depending on the input parameters and the security component, the simulation execution of our cryptosystem algorithms was on two experiments:

- In the 1st experiment: we fixed data size and varied key length in order to analyze the efficiently of the most popular asymmetric homomorphic algorithms (RSA, Paillier and ElGamal) by analyzing the process times of each algorithm's phases (encryption, computation and decryption).
- In the 2nd experiment, we fixed the key size i.e. we used the default key length recommended by NIST to preserve the security component for each algorithm, and we varied the data length.

After a successful execution, encrypted and decrypted files are created. To make sure that all the data are processed in the right way, we made a comparison between the original data file and the decrypted data file. Thus, we execute the mathematic operations performed on encrypted data for the previous cryptosystems in order to achieve our goal (efficient computations with encrypted data).

6.2 Results

6.2.1 The 1st experiment:

the simulation results of the encryption step are shown in fig.2, fig.3, table 8 and table 9, and those of the decryption step are shown in fig.4, fig.5, table 10 and table 11. The results show that the process time of RSA is less than the one of Paillier and ElGamal algorithms. For more clarification, bellow is a statistic study:

- *For small data (1KB):* the process time of the previous schemes cannot exceed a few milliseconds with a key-length less than 1024 bits. With a key >= 2048 bits, the process time of Paillier system can reach more than 4 seconds.

- *For large data (1MB):* The process time (key= 2048) of the encryption step can reach more than 42 minutes for Paillier, more than 12 minutes for ElGamal and 5 second for RSA. Concerning the process time of the decryption step, it can reach more than 72 minutes for Paillier, more than 18 minutes for ElGamal and 9 minutes for RSA.

6.2.2 The 2nd experiment

The simulation results of the encryption step are shown in table 12 and fig. 6, and those of the decryption are shown in table 13 and fig. 7. As the first execution, the results show that the time process of RSA is very less than the other Algorithms. However, with a large key, the process time take a long time to execute the steps (encryption and decryption).

6.2.3 Computations

For the mathematic operations performed on encrypted data, we multiplied/added two data cipher blocks of 32 with different key sizes:

- The simulation results of Multiplication encryption and decryption steps are shown on table12, table13, fig. 6 and fig.7.
- The additive homomorphic results for Paillier are shown in fig.8 for the encryption step and in fig.9 in the decryption step.

Table 6 The process time (ms) of the encryption / decryption of small data (1KB).

Scheme \ Key size	Encryption step				Decryption step			
	256	512	1024	2048	256	512	1024	2048
Paillier	32	74	391	2520	39	90	589	4381
RSA	4	5	43	67	20	22	95	567
ElGamal	18	40	156	1213	30	70	309	786

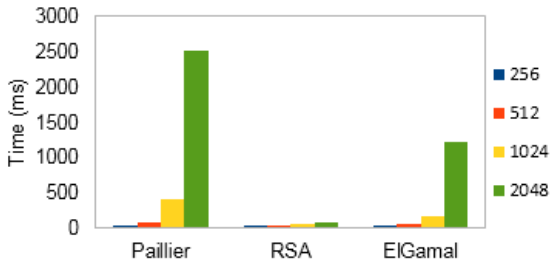


Figure 3. Encryption of small data using various schemes and different size of key

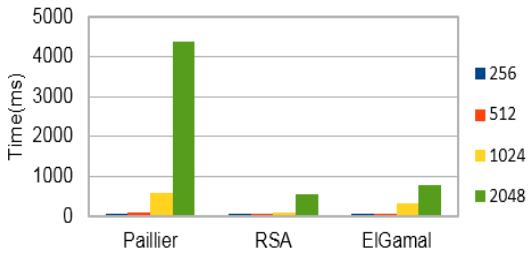


Figure 4. Decryption of small data using various scheme with different size of key

Table 7 The process time (ms) of the encryption / decryption of large data (1MB)

Key \ scheme	Encryption step				Decryption step			
	256	512	1024	2048	256	512	1024	2048
Paillier	17936	63054	236350	2633354	26030	81857	552725	4340671
RSA	711	663	1195	4855	4855	12192	76930	555541
ElGamal	12789	30082	174039	1099392	7156	15992	87118	745873

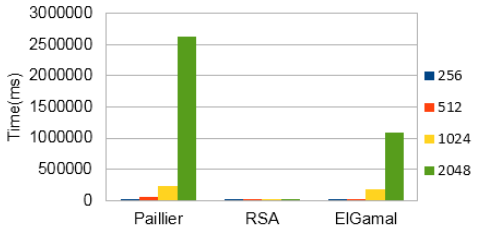


Figure 5. Encryption of large data using partially HE schemes with different sizes of key

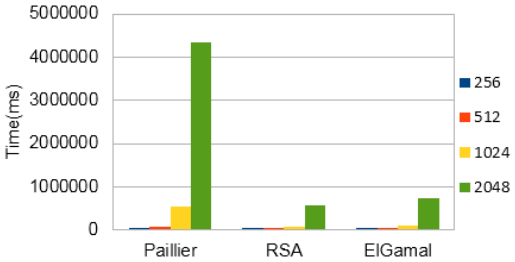


Figure 5. Decryption step of large data using various scheme with different size of key

Table 8 Time of encryption of different data size with a fixed key size (key= 2048).

Scheme \ data	32B	1KB	10KB	100KB	1MB
Paillier	79	2520	22131	243235	2633354
RSA	1	9	67	558	4855
ElGamal	43	1213	11847	109040	1099392

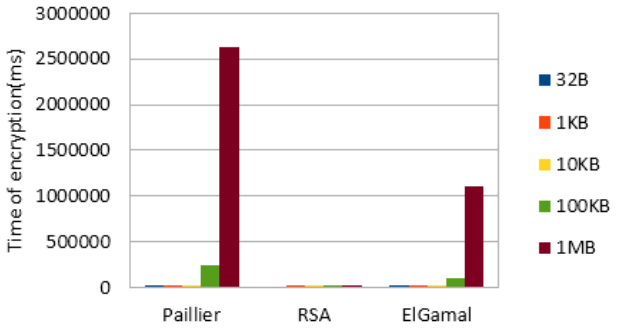


Figure 6. Time of encryption of different data size with a fixed key=2048 bits

Table 9 Decryption time of different data size with a fixed key length (key= 2048 bits).

Scheme \ data	32B	1KB	10KB	100KB	1MB
Paillier	0	4381	42129	416205	4340671
RSA	25	567	5598	56912	555541
ElGamal	19	786	6423	67532	745873

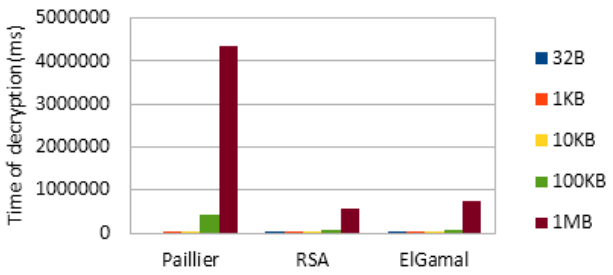


Figure 7. Time of decryption of different data sizes with a fixed key length (key= 2048)

Table 10 Time of multiplication encryption of a data block of 32B using PHE schemes.

Scheme \ Key	256	512	1024	2048
Paillier	0	0	1	4
RSA	1	1	2	2
ElGamal	1	1	1	2

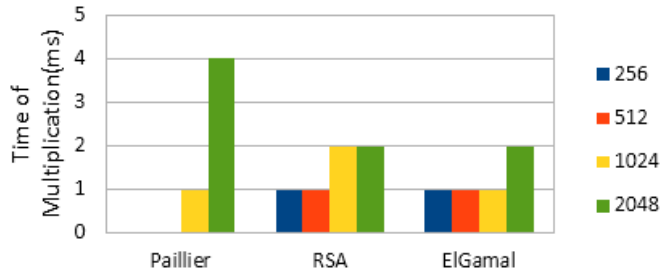


Figure 8. Time of Multiplication Decryption of a data block of 32 bytes using PHE schemes.

Table 11 Time of Multiplication Decryption of a data block of 32 bytes using PHE schemes.

Cryptosystems	256	512	1024	2048
Paillier	2	3	23	145
RSA	1	2	5	25
ElGamal	2	3	6	26

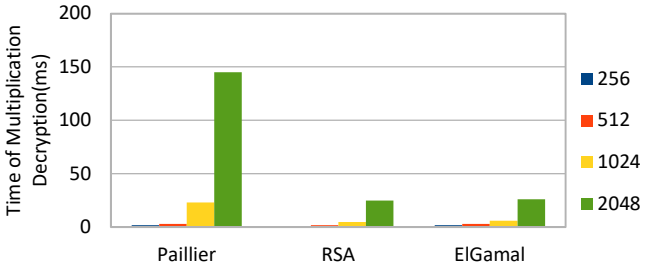


Figure 9. Time of Multiplication Decryption of a data block of 32B using PHE schemes.

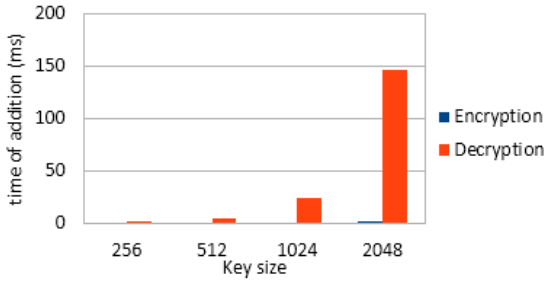


Figure 10. Addition time of encryption/decryption steps of a data block (32B) for Paillier cryptosystem.

As an analytical study of our results, RSA is faster than ElGamal and Pallier algorithms but it's less secure than the previous cryptosystems. In fact, RSA is a deterministic algorithm (we get the same cipher-text when we execute the same plaintext for many time). Contrary to ElGamal and Paillier, they are probabilistic algorithms i.e. using a random variable, you can get different cipher-texts for the same plaintext.

7 Conclusion

In this work, we described the basic concept of cloud computing and how security becomes a major issue of the delay in the widespread adoption of this technology. In addition, we presented the concept of homomorphic encryption and various encryption algorithms: additive, multiplicative and fully homomorphic encryption as well a good solution to secure data stored and computed by a third party.

This paper provides an analytic study of the most popular homomorphic encryptions schemes (RSA, Paillier and ElGamal). RSA is the faster algorithm but Paillier and ElGamal are the most efficient on term of security.

Finally, and as a reminder, our objective is how to store and enable computation on encrypted data using homomorphic encryptions proprieties in a cloud environment.

The next work is to optimize our tests of homomorphic programs by storing encrypted data on swift node and perform both of additions and multiplications operations on the cloud platform.

REFERENCES

[1] R. Chow, Ph. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Msuoka and J. Molina, “Controlling Data in the Cloud : Outsourcing Computation without Outsourcing Control,” pp. 85–90, 2009.

[2] P. V. Parmar, S. B. Padhar, S. N. Patel, N. I. Bhatt and R. H. Jhaveri, “Survey of Various Homomorphic Encryption algorithms and Schemes,” vol. 91, no. 8, p. 8887, 2014.

[3] N. Islam, W. Puech, K. Hayat and R. Brouzet, “Application of Homomorphism to Secure Image Sharing To cite this version :,” OPTICS, vol. 284, no. 19, pp. 4412–4429, 2013.

[4] M. Tebaa, S. El Hajji, and A. El Ghazi, “Homomorphic Encryption Applied to the Cloud Computing Security,” vol. I, pp. 8–11, 2012.

[5] M. Tebaa, K. Zkik, and S. El Hajji, “Hybrid Homomorphic Encryption Method for Protecting the Privacy of Banking Data in the Cloud,” vol. 9, no. 6, pp. 61–70, 2015.

[6] A. Chatterjee and I. Sengupta, “Searching and Sorting of Fully Homomorphic Encrypted Data on Cloud,” pp. 1–14.

[7] “https://github.com/hcrypt-project/libScarab” .

[8] P.Mell and T.Grance, The NIST Definition of Cloud Computing,” vol. 145, no. September, p. 6028, 2011.

[9] A. A. Atayero and O. Feyisetan, “Security Issues in Cloud Computing : The Potentials of Homomorphic Encryption,” vol. 2, no. 10, pp. 546–552, 2011.

[10] V. Vaikuntanathan, “Computing Blindfolded : New Developments in Fully Homomorphic Encryption,” 1978.

[11] S. Goldwasser, “Probabilistic Encryption *,” pp. 270–299, 1984.

[12] P. Paillier, “Public-Key Cryptosystems Based on Composite Degree Residuosity Classes,” pp. 223–238, 1999.

[13] M. Seetha and A. K. Koundinya, “Comparative Study and Performance Analysis of Encryption in RSA , ECC and Goldwasser- Micali Cryptosystems,” vol. 3, no. 1, pp. 111–118, 2014.

[14] T. ElGamal, “A public key cryptosystem and a signature scheme based on the discrete logarithm,” springer, 1985.

[15] C. Gentry, “Computing Arbitrary Functions of Encrypted Data.”

[16] C. Gentry, “A FULLY HOMOMORPHIC ENCRYPTION SCHEME,” no. September 2009.

[17] R. Meissen, “A Mathematical Approach to Fully Homomorphic Encryption ”.

[18] M. Van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, “Fully Homomorphic Encryption over the Integers,” pp. 1–28, 2010.

[19] Z. Brakerskiand C. Gentry, “Fully Homomorphic Encryption without Bootstrapping.”

[20] Z. Brakerski and V. Vaikuntanathan, “Efficient Fully Homomorphic Encryption from (Standard) LWE,” pp. 97–106, 2011.

[21] K. Lauter, M. Naehrig, and V. Vaikuntanathan, “Can Homomorphic Encryption be Practical ?,” pp. 1–18, 2011.

[22] J. W. Bos, K. Lauter, and M. Naehrig, “Private Predictive Analysis on Encrypted Medical Data.”

[23] E. Magkos, M. Burmester, and V. Chrissikopoulos, “Receipt-freeness in Large-scale Elections without Untappable Channels.”

[24] S. Iftene, “General Secret Sharing Based on the Chinese Remainder Theorem with Applications in,” Electron. Notes Theor. Comput. Sci., vol. 186, no. 3, pp. 67–84, 2007.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Comparative Study of Exact Continuous Orthogonal Moments Applications : Local Feature Extraction and Data Compression

Zaineb Bahaoui[1] , Rachid Benouini[1], Hakim EL Fadili[2], Khalid Zenkouar[3], Arsalane Zarghili[3]
[1]Laboratoire Systèmes Intelligents & Applications (LSIA), Faculty of Science Technique Fes, Université Sidi Mohamed Ben Abdellah. Fes Morocco.
[1]Laboratoire Systèmes Intelligents & Applications (LSIA), Faculty of Science Technique Fes,
[2]Laboratoire LIPI (Informatique et Physique Interdisciplinaire),ENSA School, Université Sidi Mohamed Ben Abdellah, Fes Morocco
z.bahaoui@gmail.com

ABSTRACT

This paper present an improved reconstruction algorithm of the multi-gray level images based on overlapping block method using exact continuous moments computation: Legendre , Zernike, Pseudo-Zernike and Gegenbauer moments . We solve the artifact issue caused by unitary block reconstruction which affects the visual image quality. This method aim to ensure high accuracy and low computation time, using only small finite number of moments. Our approaches aims to introduce these moments in the field of data compression and local feature extraction for pattern recognition. Experimental results show the superiority of our proposed approaches over the existing methods.

Keywords-Exact continuous moments; Legendre moments, Zernike moments; Pseudo-Zernike moments; Gegenbauer moments; Overlapping block ; Feature extraction; Compression.

1 Introduction

Since the introduction of Hu's moment methods in 1962 [1], the moments functions have been used extensively in many research fields, especially in the fields of image processing and pattern recognition, thanks to the unique global features of moment description [2-7].

As Legendre, Zernike, Pseudo-Zernike and Gegenbauer moments use the continuous orthogonal polynomials as a basis functions. They all offer a better representation capacity and increases robustness to image noise over other types of moments [2-3]. Yet they suffer from a lack of accuracy, especially due to numerical approximation errors. In order to reduce those errors, many recent works proposed a set of exact computation algorithms [8-12].

Globally, the computation of continuous moments is a time consuming procedure, mostly because of two reasons: First, the handling of a set of complicated entities for each moments order. Also, the use of great reconstruction space, involve a huge quantity of information, using high moments orders [13-15].

On the other hand, researchers have been working on establishing the mathematical formulas to describe the relationship between the global and local moment features within the same image. In this regard, we suggest a novel approach offering fast and efficient reconstruction algorithm in the case

DOI: 10.14738/tmlai.54.3226
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3226

of multi-gray level images with greater sizes, concentrating on the local feature extraction. Our main objective is to reach high reconstruction quality using only a small finite number of moments, by splitting the original image into blocks, and using fast and exact approximation. This strategy relies in the utilization of low-order polynomials on small intervals instead of high orders on a single one [13]. This aims to introduce these continues moments in the data compression field, which can be achieved by transforming the data, projecting them on the basis of functions, and then encoding the resulted coefficients [16-17]. Therefore, the input image is partitioned into blocks of pixels which are then reconstructed as separate entities.

This novel methodology based on a decomposition technique enables fast computation of image moments. Such a process is commonly used in image compression standards, e.g. in JPEG [18] where image is partitioned into blocks and DCT coefficients are computed in each part

However, when adjacent blocks have different reconstruction values, the block boundaries become visible, producing vertical and horizontal lines in the reconstructed images. This phenomenon is known as the blocking artifact [19].

To deal with this issue, we propose a new Overlapping concept, which use the neighboring information of each block and exploits the inter-block correlation. This new approach called Overlapping Block based Reconstruction (OBR), has been implemented with four different Exact continuous Moments types: Legendre (OBRELM), Zernike (OBREZM), Pseudo-Zernike(OBREPZM) and Gegenbauer (OBREGM).

In order to evaluate the performances of those four new methods, we performed a set of comparative tests on a multi-gray level images, in terms of local feature extraction either in visual quality, in PSNR or in SSIM index [20].In addition, a comparison has been conducted between a reference method, DCT, and our four proposed methods, in order to investigate whether this continuous moments an provide an alternative to the existing JPEG compression.

The rest of the paper is organized as follows. Section 2, describes the recently introduced Exact Moments computation. In section 3 we point out the theoretical details of the proposed OBRELM, OBREZM and OBREGM methods. We describe the two application of our algorithm in section 4. Finally, Section 5 gives the experimental validation with the summary of important results, and concluding remarks are presented in section 6.

2 Exactcontinuousmoments computation

2.1 Exact Legendre moments (ELM)

The $(p+q)$ order Exact Legendre Moments (ELM) [8-9], of an $M \times N$ image described by its intensity function $f(x,y)$ is:

$$L^{E}{}_{p,q} = \sum_{1}^{M}\sum_{1}^{N} I_p(x_i)\, I_q(y_j) f(x_i, y_j), \tag{1}$$

where

$$\begin{aligned} I_p(x_i) &= \tfrac{2p+1}{2}\textstyle\int_{U_i}^{U_{i+1}} P_p(x)dx/ \\ I_q(y_j) &= \tfrac{2q+1}{2}\textstyle\int_{V_j}^{V_{j+1}} P_q(y)d\ y, \end{aligned} \tag{2}$$

And

$$\int P_p(x)dx = \frac{P_{p+1}(x)-P_{p-1}(x)}{2p+1}/ \qquad (3)$$
$$\int P_q(y)dy = \frac{P_{q+1}(y)-P_{q-1}(y)}{2q+1}.$$

For the computation of Legendre polynomials, the recurrence relation can be used [8]:

$$P_{p+1}(x) = \frac{(2p+1)}{(p+1)}xP_p(x) - \frac{P}{(p+1)}P_{p-1}(x). \qquad (4)$$

For simplicity, upper and lower limits of the integration in (2) and (3) will be expressed as follows:

$$U_{i+1} = -1 + i\Delta x, U_i = -1 + (i-1)\Delta x. \qquad (5)$$

Similarly,

$$V_{j+1} = -1 + j\Delta y, V_j = -1 + (j-1)\Delta y. \qquad (6)$$

Using (2), (3) and (4), the integral parts will be written as follows:

$$\int_{U_i}^{U_{i+1}} P_p(x)dx = \left[\frac{P_{p+1}(x)-P_{p-1}(x)}{2p+1}\right]_{U_i}^{U_{i+1}}/ \qquad (7)$$
$$\int_{V_j}^{V_{j+1}} P_q(y)dx = \left[\frac{P_{q+1}(y)-P_{q-1}(y)}{2q+1}\right]_{V_j}^{V_{j+1}}.$$

Substitute $P_{p+1}(x)$ from (4) into (7) yields (8):

$$\int_{U_i}^{U_{i+1}} P_p(x)dx = \frac{1}{p+1}\left[xP_{p+1}(x) - P_{p-1}(x)\right]_{U_i}^{U_{i+1}} / \int_{V_j}^{V_{j+1}} P_q(y)dy = \frac{1}{q+1}\left[yP_{q+1}(y) - P_{q-1}(y)\right]_{V_j}^{V_{j+1}}. \qquad (8)$$

The set of Legendre moments can thus be computed exactly by (1) where

$$I_p(x_i) = \frac{2p+1}{2p+2}\left[xP_{p+1}(x) - P_{p-1}(x)\right]_{U_i}^{U_{i+1}} /$$
$$I_q(y_j) = \frac{2q+1}{2q+2}\left[yP_{q+1}(y) - P_{q-1}(y)\right]_{V_j}^{V_{j+1}}. \qquad (9)$$

Equation (1) is called the exact computation of Legendre moments (ELM) [8-9].

The image function $f(x,y)$ can be written as an infinite series of expansion in terms of the Legendre polynomials over the square $[-1,1]$

$$f(x,y) = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} L_{p,q}\ P_p(x)\,P_q(y), \qquad (10)$$

where the Legendre moments $L^E_{p,q}$ are computed over the same square. If only Legendre moments of order $p+q \leq max$ are given, then the function $f(x,y)$ can be approximated by a truncated finite series :

$$\hat{f}_{max}(x,y) = \sum_{p=0}^{max}\sum_{q=0}^{p} L_{p-q,q}P_{p-q}(x)P_q(y), \qquad (11)$$

where the number of moments used in this form for image reconstruction is defined by [3]:

$$N_{total} = \frac{(max+1)(max+2)}{2}. \qquad (12)$$

2.2 Exact Zernike moments (EZM)

Exact Zernike moments are computed by using exact geometric moments, proposed by Hosny in [10]. The approach by first calculating the geometric moments accurately, then employing the relationship in (13)of Zernike moments and the geometric moments.

$$Z_{pq} = \frac{p+1}{\pi} \sum_{\substack{k=|q| \\ p-k=even}}^{p} \sum_{j=0}^{s} \sum_{m=0}^{|q|} w^m \binom{s}{j} \binom{|q|}{m} B_{p|q|k} G_{k-2j-m,\ 2j+m}, \tag{13}$$

where$s = 0.5(k - |q|)$ and $\hat{\imath} = \sqrt{-1}$

$$w = \begin{cases} -\hat{\imath}, & q > 0 \\ \hat{\imath}, & q \leq 0 \end{cases}, \tag{14}$$

The Exact Zernike moments$Z_{pq}^{E}$is simplified as following:

$$Z_{pq}^{E} = \frac{p+1}{\pi} \sum_{\substack{k=|q| \\ p-k=even}}^{p-2} B_{p|q|k} Rad_{kq} \tag{15}$$

Zernike polynomial coefficients are computed using the following recurrence relations:

$$B_{ppp} = 1, \tag{16}$$

$$B_{p(q-2)p} = \frac{p+q}{p-q+2} B_{pqp}, \tag{17}$$

$$B_{pq(k-2)} = \frac{(k+q)(k-q)}{(p+k)(p-k+2)} B_{pqk}. \tag{18}$$

The time consuming direct computations of factorial terms are avoided by using the following recurrence relations:

$$D(p,k) = \frac{p+q}{p-k} D(p-1,k), \tag{19}$$

$$D(p,k) = \frac{1}{k(p-k)} D(p,k-1), \tag{20}$$

with$D(0,0) = 1$,and $D(p,0) = 1$.

The radial Zernike moments are expressed as a combination of geometric moments.

$$Rad_{kq} = \sum_{j=0}^{s} \sum_{m=0}^{|q|} w^m D(s,j) D(q,m) G_{k-2j-m,\ 2j+m}. \tag{21}$$

Hosny in [46] proposed an exact and fast geometric moments's computation. In this exact method, the set of geometric moments can thus be computed exactly by:

$$G_{pq} = \sum_{i=0}^{N} \sum_{j=0}^{N} I_p(i) I_q(j)\, f(x_i, y_j), \tag{22}$$

where

$$I_p(i) = \frac{1}{p+1} \left[U_{i+1}^{p+1} - U_i^{p+1} \right] /$$
$$I_q(j) = \frac{1}{q+1} \left[U_{j+1}^{q+1} - U_j^{q+1} \right], \tag{23}$$

Fast computation of exact geometric moments can be achieved by successive computation of the qth order moments for each row. (22)will be rewritten in a separable form as follows:

$$G_{pq} = \sum_{i=0}^{N} I_p(i) Y_{iq}(j). \tag{24}$$

where

$$Y_{iq} = \sum_{j=0}^{N} I_q(j) f(x_i, y_j). \tag{25}$$

Y_{iq}in(25)is the q^{th} order moments of row i. Since,$I_0(i) = \frac{\sqrt{2}}{N}$.

Substitute $I_0(i)$ into (23) yields;

$$G_{0q} = \frac{\sqrt{2}}{N} \sum_{i=0}^{N} Y_{iq}(j). \tag{26}$$

The image intensity function $f(x, y)$ can then be expressed in terms of Zernike polynomials over the unit circle as

$$f(x,y) = \sum_{p=0}^{\infty} \sum_{\substack{q \\ p-q=even \\ q \le p}} \lambda_p Z_{pq} V_{pq}(x,y), \tag{27}$$

whereλ_p is a normalization constant that depends on the square to circular mapping technique and

$$V_{pq}(x,y) = R_{pq}(r) e^{-iq\theta}, r \in [-1,1] \tag{28}$$

Where $r = \sqrt{x^2 + y^2}$ is the length of the vector from the origin to the pixel (x, y), and $\theta = \tan^{-1}(y/x)$ is the angle between the vector r and the principle x-axis. The real valued radial polynomials are given by

$$R_{pq}(r) = \sum_{\substack{k=q \\ p-k=even}}^{p} B_{p|q|k} r^k, \tag{29}$$

where the polynomial coefficients, $B_{p|q|k}$ are defined as:

$$B_{p|q|k} = \frac{(-1)^{\left(\frac{p-k}{2}\right)} \left(\frac{p+k}{2}\right)!}{\left(\frac{p-k}{2}\right)! \left(\frac{k+q}{2}\right)! \left(\frac{k-q}{2}\right)!}, \tag{30}$$

With Zernike moments,Z_{pq} are computed over the same unit circle. However, the reconstruction of image function using arbitrary large order of Zernike moments is not efficient. Therefore, this series expansion is truncated at a finite order, Max, and considered as optimum approximation to the original image function. The reconstructed image intensity function$\hat{f}(x, y)$, based on the truncated series, is given as:

$$\hat{f}_{Max}(x,y) \approx \sum_{p=0}^{Max} \sum_{\substack{q \\ p-q=even \\ q \le p}}^{p} \lambda_p Z_{pq} V_{pq}(x,y). \tag{31}$$

Where the number of moments used in this form for image reconstruction is definedin [3] by:

$$N_{total} = \begin{cases} (\frac{Max+2}{2})^2, & Maxiseven \\ (\frac{Max+1}{2})^2 + \frac{Max+1}{2}, & Maxisodd \end{cases} \quad (32)$$

2.3 Exact Pseudo-Zernike moments (EPZM)

EPZM are computed as a linear combination of exact geometric and radial geometric moments, where they are exactly computed by using mathematical integration [11].

$$PZ_{pq} = \frac{p+1}{\pi} \left(\sum_{\substack{k=q \\ k-q=even}}^{p} B_{p;q;k} R^{G}_{k,q} + \sum_{\substack{k=q+1 \\ k-q=odd}}^{p} B_{p;q;k} R^{H}_{k,q} \right), \quad (33)$$

where

$$R^{G}_{k,q} = \sum_{j=0}^{s} \sum_{m=0}^{q} (-\hat{\imath})^m \, C_j^S C_m^q \, GM_{k-2j-m,\ 2j+m}, \quad (34)$$

$$R^{H}_{k,q} = \sum_{j=0}^{s} \sum_{m=0}^{q} (-\hat{\imath})^m \, C_j^S C_m^q \, H_{k-2j-m,\ 2j+m}. \quad (35)$$

The PZM coefficients B_{pqk} are computed using the following recurrence relations:

$$B_{ppp} = 1, \quad (36)$$

$$B_{p(q-1)p} = \frac{k+p+1}{p-q+1} B_{pqp}, \quad (37)$$

$$B_{pq(k-1)} = \frac{(k+q+1)(k-q)}{(p+k+1)(p-k+1)} B_{pqk}. \quad (38)$$

The geometric and radial geometric moments GM_{pq} and H_{pq} are defined in function, of image intensity $f(x, y)$, as follows:

$$GM_{pq} = \sum_{i=0}^{[\frac{N}{2}]} \sum_{j=0}^{[\frac{N}{2}]} I_p(i) I_q(j) \, f(x_i, y_j), \quad (39)$$

$$H_{pq} = \sum_{i=0}^{[\frac{N}{2}]} \sum_{j=0}^{[\frac{N}{2}]} I_p(i) I_q(j) \sqrt{x^2 + y^2} f(x_i, y_j), \quad (40)$$

where $I_p(x)$ and $I_q(y)$ are calculated by Eq.(23).

Where the number of moments used in this form for image reconstruction is defined in [3] by:

$$N_{total} = (Max + 1)^2 \quad (41)$$

2.4 Exact Gegenbauer moments (EGM)

The accurate orthogonal Gegenbauer moments of order *(p, q)* [12] with the scaling parameter $\alpha > -0.5$ are defined as follows:

$$G^{(\alpha)}_{p,q} = \frac{1}{C_p(\alpha) C_q(\alpha)} \sum_{1}^{M} \sum_{1}^{N} IX_p(x_i) \, IY_q(y_j) f(x_i, y_j), \quad (42)$$

where the normalization constant $C_p(\alpha)$ defined with the recurrence relation as:

$$C_p(\alpha) = \frac{(p-1+\alpha)(p-1+2\alpha)}{p(p+\alpha)} C_{p-1}(\alpha) \tag{43}$$

$$C_0(\alpha) = \frac{\pi\Gamma(2\alpha)}{2^{2\alpha-1}\alpha[\Gamma(\alpha)]^2}$$

The symbol, Γ(.), refers to gamma function and:

$$IX_p(x_i) = \sum_{k=0}^{[\frac{p}{2}]} B_{p,k}^{(\alpha)} I_{p-2k}(x_i) \tag{45}$$

$$IY_q(y_j) = \sum_{k=0}^{[\frac{q}{2}]} B_{q,k}^{(\alpha)} I_{q-2k}(y_j) \tag{46}$$

The floor operator $[p/2]$, equal to either $(p-1)/2$ or $p/2$ for odd and even values of p, and the coefficients matrix $B_{p,k}^{(\alpha)}$ obey the following recursive relations

$$B_{0,0}^{(\alpha)} = 1, \tag{47}$$

$$B_{p,0}^{(\alpha)} = \frac{2(\alpha+p-1)}{p} B_{p-1,0}^{(\alpha)}, \tag{48}$$

$$B_{p,k}^{(\alpha)} = \frac{(p-2k+2)(p-2k+1)}{4k(\alpha+p-k)} B_{p,k-1}^{(\alpha)}, \tag{49}$$

The definite integration $I_{p-2k}(x_i)$and $I_{q-2k}(y_j)$ is evaluated using the recurrence relation:

$$I_T = \frac{1}{(2\alpha+T)}[U_i^{T-1}(1-U_i^2)^{\alpha+0.5} - U_{i+1}^{T-1}(1-U_{i+1}^2)^{\alpha+0.5} + (T-1)I_{T-2} \tag{50}$$

where$T$ is a non-negative integer, $T = 2, 3, 4, \ldots..Max$,with Max is the maximum order of Gegenbauer moments, and $I_0 = \int_{U_i}^{U_{i+1}} (1-x^2)^{\alpha-0.5} dx$ and $I_1 = \frac{1}{(2\alpha+1)}[(1-U_i^2)^{\alpha+0.5} - (1-U_{i+1}^2)^{\alpha+0.5}]$.

The integration of I_0 is difficult to be evaluated analytically. So, an accurate numerical integration method is a good choice. The composite Simpson's rule is proved to be very accurate where the numerical and exact values are almost equal [12].

For the computation of Gegenbauer polynomials, the recurrence relation can be used :

$$P_{p+1}^{(\alpha)}(x) = \frac{(2p+2\alpha)}{(p+2\alpha)} x P_p^{(\alpha)}(x) - \frac{p+2\alpha-1}{(p+1)} P_{p-1}^{(\alpha)}(x) \tag{51}$$

with$P_0^{(\alpha)}(x) = 1$, $P_1^{(\alpha)}(x) = 2\alpha x$ and $p \geq 0$.

The image function $f(x,y)$ can be written as an infinite expansion series in terms of the Gegenbauer polynomials over the square$[-1,1]x\ [-1,1]$:

$$f(x,y) = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} G_{p,q}^{(\alpha)} P_p^{(\alpha)}(x) P_q^{(\alpha)}(y) \tag{52}$$

where the Gegenbauer moments, $G_{p,q}^{(\alpha)}$, are computed over the same square. If only Gegenbauer moments of order smaller than or equalto Max are given, then the function $f(x,y)$ in equation (43) can bereconstructed as follows:

$$\hat{f}_{Max}(x,y) \approx \sum_{p=0}^{Max} \sum_{q=0}^{p} G_{p-q,q}^{(\alpha)} P_{p-q}^{(\alpha)}(x) P_{q}^{(\alpha)}(y) \tag{53}$$

Where the number of moments used in this form for image reconstruction is defined in Eq.(12).

3 Overlapping blocks reconstruction using Exact moments

We learnt from the previous chapter, that regardless of the moments used, It is established that the exact momentscomputation can reduce significantly the reconstruction error rate and increase the outcome quality.

Nevertheless, if we consider images with greater sizes, higher order moments are involved and the handling of greater quantity of information is necessary [13]. Hence the moments computation becomes a time consuming procedure.

Nevertheless, if we consider images with greater sizes, higher order moments are involved and the handling of greater quantity of information is necessary [13]. Hence the moments computation becomes a time consuming procedure.

To overcome this limitation, we first proposed an approach using block based reconstruction method: The input image is subdivided into square blocks of pixels of size (k,l),which generates a number of sub-images reconstructed separately. Because sub-image size is smaller, only low moments's order are needed to better describe its content. Thus, this method allows faster computation of image moments.

However, since each block is handled as an independent entity, discontinuities occur at the block boundaries. This is known as the blocking artifact, which is a consequence of the lack of inter block correlation during the reconstruction process.

In order to exploits this inter block correlation, our proposed approach relies on the overlapping concept: We use for each block the three adjacent neighboring information's by performing the moments computation process on the Overlapped blocks and reconstructing the initial non-overlapped block. In our case, the blocking artifact is dealt with at the source, namely, the moments computation process.

We applied this Overlapping approach to each of the four moments types: Legendre, Zernike, Pseudo-Zernike and Gegenbauer. Hence, we will propose four reconstruction methods:

OBRELM: Overlapping Block based Reconstruction using Exact Legendre Moments.

OBREZM: Overlapping Block based Reconstruction using Exact Zernike Moments.

OBREPZM: Overlapping Block based Reconstruction using Exact Pseudo- Zernike Moments.

OBREGM: Overlapping Block based Reconstruction using Exact Gegenbauer Moments.

Those methods are performed through two stages: (i) the exact moments computation which extracts the block neighborhood information by proceeding on overlapped blocks; (ii) the reconstruction process which acts on output blocks and merge them into the final image (Figure. 1).

The following figures show the reconstruction stages and the Overlapping scheme.

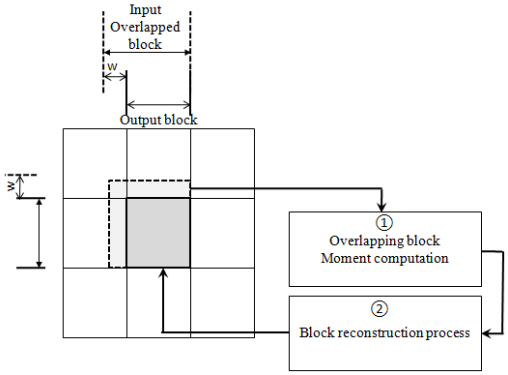


Figure 1 The overlapping scheme of Legendre and Gegenbauer moments, which are defined in the unit square

As we can see in Figure. 1 this overlapping scheme uses the information of the tree neighbors block, It's a one side overlapping, the other side will be overlapped with the next block. In the case of Zernike and Pseudo-Zernikemoments, our overlapping method also proceed by partionning the input image into square blocks of size (k, l), which are then mapped on circular blocks (r, θ) parameters [10].

We consider w extra pixel on one side (Figure.1), the pixels of the additional perimeter would be considered in the moment's computation of this block.

Two works [15] and [19], have proved that an overlapping window size of one pixel is the most effective in terms of image quality and time consuming.

By using the overlapping method, the image space will take the following form:

$$\Omega_{\text{Overlap}} = \{x_i, y_j | 0 \le x_i \le M - 1 + w, \qquad 0 \le y_j \le N - 1 + w\} \tag{53}$$

We assume that:

$$\Omega_{\text{Overlap}} = \bigcup_{n_1=0}^{(S_1-1)} \bigcup_{n_2=0}^{(S_2-1)} D_{\text{Overlap}}^{n_1,n_2}. \tag{54}$$

Where the subset $D_{\text{Overlap}}^{n_1,n_2}$ is expressed as:

$$D_{\text{Overlap}}^{n_1,n_2} = \{x_i, y_j | n_2 k \le x_i \le (n_2 + 1)k - 1 + w, n_1 l \le y_j \le (n_1 + 1)l - 1 + w \ \}. \tag{55}$$

Then let the image function associated to each $D_{\text{Overlap}}^{n_1,n_2}$subset be defined as follows:

$$f_{\text{Overlap}}^{n_1,n_2}(x, y) = \left\{ f(x_i, y_j) \middle| x_i, y_j \in D_{\text{0verlap}}^{n_1,n_2} \right\}. \tag{56}$$

This gives:

$$f(x, y) = \bigcup_{n_1=0}^{(S_1-1)} \bigcup_{n_2=0}^{(S_2-1)} f_{\text{Overlap}}^{n_1,n_2}(x, y) \ . \tag{57}$$

We introduce The Legendre moments defined on the subspace $D_{\text{Overlap}}^{n_1,n_2}$for the Overlapping Block based Reconstruction using Exact Legendre Moments method (OBRELM) is:

$$L_{p,q,Olap}^{E,n_1,n_2} = \sum_{i=n_2k}^{(n_2+1)k-1+w} \sum_{j=n_1l}^{(n_1+1)l-1+w} I_{p,Olap}^{n_1}(x_i)\, I_{q,Olap}^{n_2}(y_j) f^{n_1,n_2}(x_i, y_j) \tag{58}$$

The Zernike moments defined for the Overlapping Block based Reconstruction using Exact Zernike Moments method (OBREZM) is:

$$Z_{pq,Overlap}^{E,n_1,n_2} = \frac{p+1}{\pi} \sum_{\substack{k=|q| \\ p-k=even}}^{p-2} B_{p|q|k} Rad_{kq,Overlap}^{n_1,n_2}. \tag{59}$$

The Pseudo-Zernike moments defined for the Overlapping Block based Reconstruction using Exact Zernike Moments method (OBREPZM) is:

$$PZ_{pq,Ovlap}^{n_1,n_2} = \frac{p+1}{\pi} \left(\sum_{\substack{k=q \\ k-q=even}}^{p} B_{p;q;k} R_{k,q}^{G,n_1,n_2} + \sum_{\substack{k=q+1 \\ k-q=odd}}^{p} B_{p;q;k} R_{k,q}^{H,n_1,n_2} \right),. \tag{60}$$

The Gegenbauer moments defined for the Overlapping Block based Reconstruction using Exact Gegenbauer Moments method (OBREGM) is:

$$G_{p,q,Ovlap}^{E,n_1,n_2} = \frac{1}{C_p(\alpha) C_q(\alpha)} \sum_{i=n_2k}^{(n_2+1)k-1+w} \sum_{j=n_1l}^{(n_1+1)l-1+w} IX_{p,Olap}^{n_1}(x_i)\, IY_{q,Olap}^{n_2}(y_j) f^{n_1,n_2}(x_i, y_j). \tag{61}$$

Then, the functions of each image block using the OBRELM, OBREZM, OBREPZM and OBREGM are as follows:

$$\hat{f}_{max,Ovlap}^{n_1,n_2}(x_i, y_j) \approx \sum_{p=0}^{max} \sum_{q=0}^{p} L_{p-q,q,Ovlap}^{E,n_1,n_2} P_{p-q,Ovlap}(x_i) P_{q,Ovlap}(y_j). \tag{62}$$

$$\hat{f}_{max,Ovlap}^{n_1,n_2}(x_i, y_j) \approx \sum_{p=0}^{max} \sum_{\substack{q \\ p-q=even \\ q \le p}}^{p} \lambda_p Z_{p,q,Ovlap}^{E,n1,n2} V_{p,q}^{n1,n2}(x_i, y_j). \tag{63}$$

$$\hat{f}_{max,Ovlap}^{n_1,n_2}(x_i, y_j) \approx \sum_{p=0}^{max} \sum_{\substack{q \\ p-q=even \\ q \le p}}^{p} \lambda_p Z_{p,q,Ovlap}^{E,n1,n2} V_{p,q}^{n1,n2}(x_i, y_j). \tag{64}$$

$$\hat{f}^{n_1,n_2}_{max,Ovlap}(x_i,y_j) \approx$$

$$\sum_{p=0}^{max}\sum_{q=0}^{p} G^{E,n_1,n_2}_{p-q,q,Ovlap} P_{p-q,Ovlap}(x_i) P_{q,Ovlap}(y_j).$$

The image function up to max of the three approaches using Legendre, Zernike and Gegenbauer moments can be finally obtained by:

$$\hat{f}_{max}(x,y) = \bigcup_{n_1=0}^{(S_1-1)} \bigcup_{n_2=0}^{(S_2-1)} \hat{f}^{n_1,n_2}_{max,Overlap}(x,y), \quad (66)$$

Our approaches OBRELM, OBREZM, OBREPZM and OBREGM using respectively Legendre, Zernike, Pseudo-Zernike and Gegenbauer moments achieve improvement in the following points:

i. Giving high reconstruction quality by using only a small finite number of moments.
ii. Mitigating the artifact involved in the block processing by exploiting the block neighborhood information during the moments computation step, which allows to avoid enhancement post processing techniques which are a time-consuming procedures.

This property can be used to extract local features from the desired Region of interest (ROI). Also, it can be applied in the field of data compression, in which we aim to reconstruct the original image using only a finite number of moments.

4 Applications of our four proposed methodes

4.1 Local features extraction

In this section, we show that we can use our proposed block representation approaches, based on continuous moments, for local features extraction from an image. The local features can be extracted easily from any desired location in the image (region of interest), due to the capability of our approaches to separately represent the information of each block. Consequently, their use in object classification and recognition applications is highly significant. Our proposed local features extraction algorithm is described as follows: firstly; the input image is divided into small blocks. Secondly, for each block that correspond to the desired ROI, we apply our proposed methods (OBRELM, OBREZM, OBREPZM and OBREGM) for extracting the local features. Finally, we transform the resulting moments coefficients to a classifier (Figure 2).

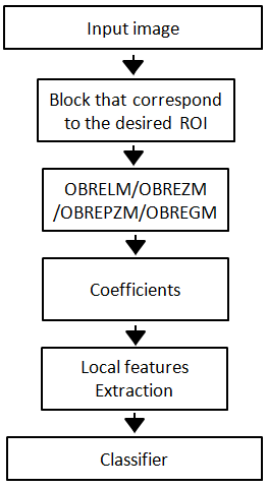


Figure 1 The principe of applying our four approaches in local feature extraction for classification

4.2 Compression algorithm

The compression algorithm is described as follows: Each images is first divided into sub-blocks whose size is (M x M), each block is then transformed by using the OBRELM, OBREZM, OBREPZM and OBREGM. At last, the set of calculated moments are transmitted to the receiver to be used in the reconstruction step (Figure .3) .

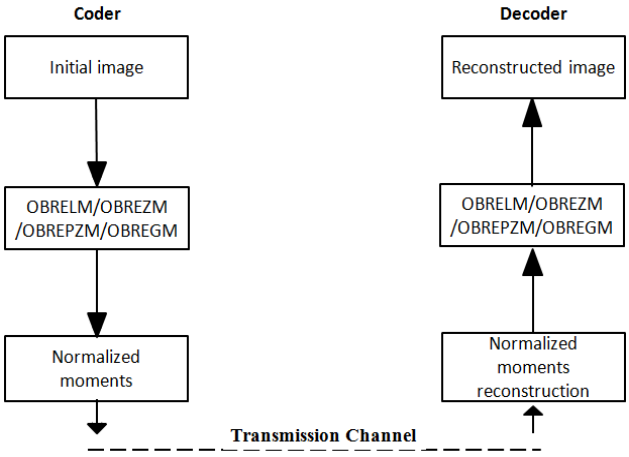


Figure 2 Diagram of encoding/ decoding scheme

In order to evaluate the performance of different methods, we use the total compression ratio [17] which is defined as follows:

$$TCR = \frac{image\ debit}{number\ of\ moments} \tag{67}$$

5 Experimental results

This section aims to prove the effectiveness of the proposed methods using Legendre, Zernike, pseudo-Zernike and Gegenbauer moments in terms of compression rate and local features extraction. Therefore, a (128 x 128) real gray level "clock" test image is used to compare our four proposed methods in terms of local features extraction. In order to complete our study, a comparison of our approaches with DCT for image compression is made in terms of SSIM index, PSNR and time consuming using 128 x 128 Lena image.

5.1 Local features extraction

To measure the capacity of our method, using continuous orthogonal moments, in term of local feature extraction, we have conducted the local sub-image reconstruction via our approaches on a Clock image. Firstly, we reconstruct just the clock object with only small moments orders. Then, we have adopted the Peak Signal-to-Noise ratio (PSNR) and SSIM index [20] as the measurements to evaluate the reconstructed images.

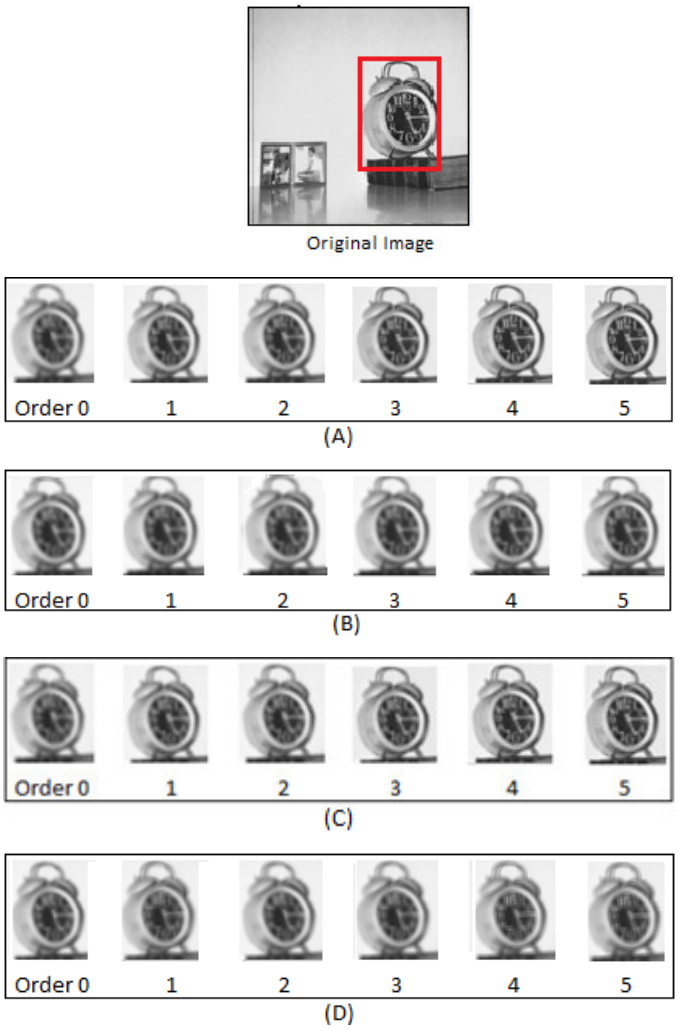


Figure 3 Local Reconstuction image using Overlapping block approaches of Clock image using Legendre moments (OBRELM) (A), Zernike moments (OBREZM) (B),), Pseudo-Zernike moments (OBREPZM) (C) and Gegenbauer moments (OBREGM) (D) with different orders

Tablem 1 Values of the PSNR/SSIM for the local reconstructed clock images using the methods based on overlapping block: OBRELM, OBREZM, OBREPZM and OBREGM using respectively Legendre, Zernike,Pseudo-Zernike and Gegenbauer moments with block size (4x4).

	OBRELM	OBREZM	OBREPZM	OBREGM
Order	PSNR/SSIM			
0	23.55/0.89	23.15/0.88	23.50/0.90	20.77/0.79
1	25.38/0.91	23.48/0.89	25.79/0.90	21.30/0.83
2	27.73/0.94	25.24/0.90	28.02/0.94	22.34/0.89
3	30.05/0.95	27.40/0.92	30.76/0.96	24.39/0.91
4	32.46/0.97	30.95/0.96	33.13/0.98	28.85/0.92
5	33.35/0.99	31.08/0.98	33.56/0.99	30.04/0.94

The above results (Figure.4, Table.1) illustrates the efficiency of the proposed methods against the blocking artifact. They show how a relatively small finite set of moments can adequately characterize the given image with no need to include higher order moments. The OBRELM and OBREPZM perform the best, OBREZM using Zernike moments performs well throughout the reconstruction process and the OBREGM using Gegenbauer moments perform lower than the others.

5.2 Compression: Comparison with DCT

In this section, comparison is made between our approaches and DCT for image compression. The information is stored in some computed moments and coefficients in our three methods and DCT respectively. After reconstruction, the PSNR and SSIM produces by the four methods are compared.

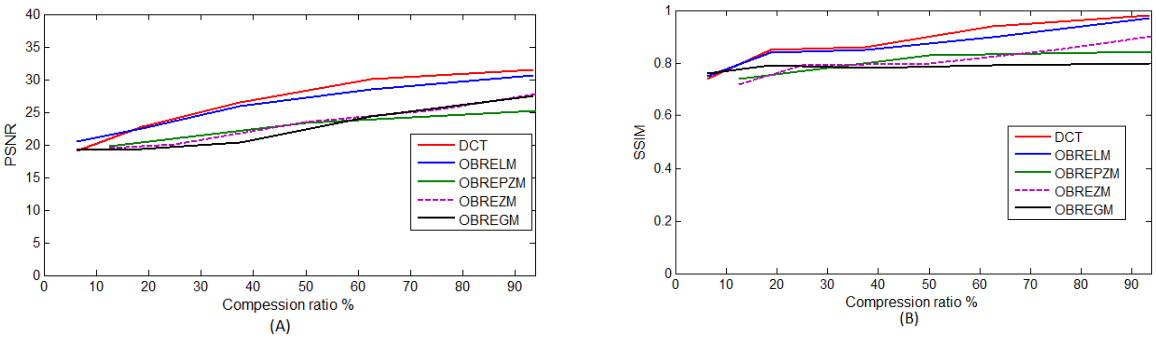


Figure 4 Comparison between OBRELM, OBREZM, OBREGM and DCT for (128 x 128) Lena image using block size (4 x 4) in terms of PSNR (a) and SSIM Index (b)

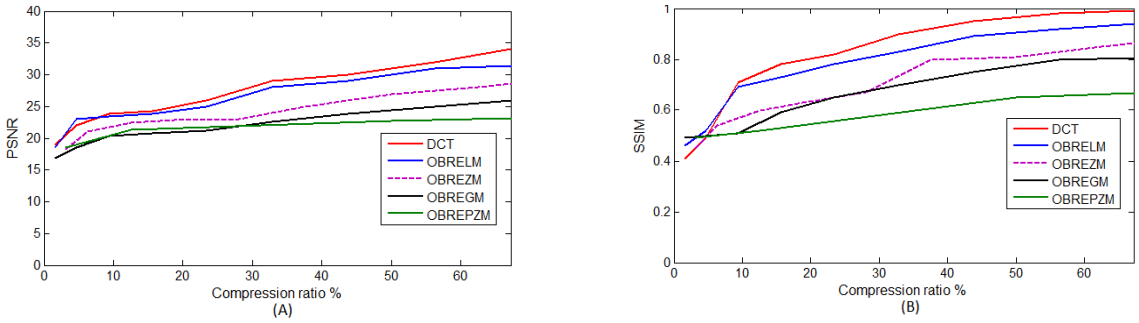


Figure 5 comparison between OBRELM, OBREZM, OBREGM and DCT for (128 x 128) Lena image using block size (8 x 8) in terms of PSNR (a) and SSIM Index (b)

Table 3The reduction factors of reconstruction time for our proposed methods, using 128 x 128 Lena image and (4 x 4) block size, for PSNR = 20, compared to the global reconstruction with 2.0 GHz i7 and 8 GB RAM.

Method	**OBRELM**	**OBREZM**	**OBREPZM**	**OBREGM**
Corresponding moments Order	0	1	0	2
CPU elapsed time	0.064s	1.87s	12.10s	0.58s
Time Reduction	97.58%	92.23%	75.56%	81.62%

The results in Figure. 5 and Figure. 6 indicates that reconstruction using Legendre moments OBRELM provide a better compression capability compared to Zernike, Pseudo-Zernike and Gegenbauer moments which is. Whereas the OBRELM is slightly lower than DCT. However, for high compression rates the performance is almost similar.

For a better compression ratio, a quantification and numerical redundancy exploitation and entropy coding method such as Huffman coding [27] can be used.

Table 2 confirms that the OBRELM is a way better in term of data compression, but also in time reduction. This also means the other methods are unsuitable because the compression schemes require very fast encoding/ decoding image with low time consumption.

The OBRELM is better than the OBREZM and OBREPZM due to the fact that Zernike and Pseudo-Zernike moments are a complex numbers and require, for a given order, two parts in the computation: magnitude and phase. Hence, in the step of Zernike and Pseudo-Zernike moments reconstruction it will need an additional computation of the two values and, by the way, increases the complexity of the overall process. Some works used only magnitude information for recognition but involves erroneous results and impreciseness [28].

6 Conclusion

In this paper, a novel and faster algorithm for the computation of exact continuous moments: Legendre, Zernike, Pseudo-Zernike and Gegenbauer, have been presented. We proved that, by replacing a greater size gray level image by a set of blocks, the moments computation can be accelerated significantly by using only small finite number of moments. We have also applied those methods to local features extraction and to data compression. The experimental results prove that the proposed methods outperforms the conventional ones in terms of error reduction, image quality and time consumption. Our OBREPZM method seems to be the best approach in term of local features extraction but it is high time consuming. In the other hand, our OBRELM method is the most suitable for data compression. Indeed, our approach allows a high compression ratio and deliver a good visual quality and low reconstruction error with less time consuming.

REFERENCES

[1] M. K. Hu, “Visual problem recognition by moment invariant,” IRE Trans. Inform. Theory, vol. IT-8, pp. 179-187, Feb. 1962.

[2] M. Teague, “Image analysis via the general theory of moments,” J. Opt. Soc. Amer., vol. 70, no. 8, pp. 920-930, 1980

[3] C.-H. Teh and R. T. Chin, “On image analysis by the methods of moments,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 10, no. 4, pp. 496–513, 1988.

[4] Krawtchook liao

[5] J. Haddadnia, K. Faez, and P. Moallem, “Neural network based face recognition with moment invariants,” Proceedings of ICIP2001, Thessaloniki, Greece, pp. 1018-1021, 2001.

[6] A. Khotanzad, Y.H. Hong, "Invariant image recognition by Zernike moments", IEEETrans. Pattern Anal. Mach. Intell. 12 (1990) 489–497

[7] R. Luo, T. Lin, Finger crease pattern recognition using Legendre moments and principal component analysis, Chin. Opt. Lett. 5(3), 160–163 (2007).

[8] P.-T. Yap, R. Paramesran, “An efficient method for the computation of Legendre moments,” IEEE Trans. Pattern Anal. Mach. Intell, vol. 27, no. 12, pp. 1996–2002,2005

[9] K.M. Hosny, “Exact Legendre moment computation for gray level images,” Pattern Recognition, vol. 40, no. 12, pp. 3597–3605, 2007.

[10] K.M. Hosny, "Fast computation of accurate Zernike moments", Journal of Real-Time Image Processing 3 (2008) 97–107.

[11] K.M. Hosny, K.M. "Fast computation of accurate Pseudo-Zernike moments for binary and gray-level images". Int Arab J Inf Technol., 2014, 11, (3).

[12] K.M. Hosny, "Image representation using accurate orthogonal Gegenbauer moments". Pattern Recognit. Lett. 2011, 32, (6), pp. 795–804

[13] H. El Fadili, K. Zenkouar and H. Qjidaa, “Lapped block image analysis via the method of Legendre moments,” EURASIP Journal on Applied Signal Processing, vol. 2003, no.9, pp. 902-913, 2003.

[14] Z. Bahaoui, K. Zenkouar, H. El fadili, H.Qjidaa, A.Zarghili, Global overlapping block based reconstruction using exact Legendre moments, CIST (2014), pp. 323 - 328.

[15] Z. Bahaoui, K. Zenkouar, H. El fadili, H.Qjidaa, A.Zarghili, blocking artifact removal using partial overlapping based on exact Legendre moments computation, Journal of Real-Time Image Processing, (DOI :10.1007/s11554-014-0465-3) (2014)1861-8200.

[16] S.G. Mallat, A theory for multiresolution signal decomposition the wavelet representation, IEEE Trans. Pattern Anal. Machine Intel. 11 (7) 1989.

[17] H. Zenkouar, A. Nachit, "Images compression using moments method of orthogonal polynomials", Materials Science and Engineering: B, 49(3), pp. 211–215 (1997)

[18] ISO/IEC 10918-1, Digital compression and coding of continuous-tone still images: requirements and guidelines, February, 1994.

[19] Reeve III, H.C., Lim, J.S.: Reduction of blocking effect in image coding, Proceedings of ICASSP 83, 1212 -1215 (1988).

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” Image Processing, IEEE Transactions on, 2004, vol. 13, no. 4, pp. 600–612

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Data Editing for Semi-Supervised Co-Forest by the Local Cut Edge Weight Statistic Graph (CEWS-Co-Forest)

Nesma Settouti, Mohammed El Amine Bechar, Mostafa EL Habib Daho, Mohammed Amine Chikh
Biomedical Engineering Laboratory, Tlemcen University, Algeria
acmlis.conference@gmail.com

ABSTRACT

In order to address the large amount of unlabeled training data problem, many semi-supervised algorithms have been proposed. The training data in semi-supervised learning may contain much noise due to the insufficient number of labeled data in training set. Such noise may snowball themselves in the following learning process and thus hurt the generalization ability of the final hypothesis. If such noise could be identified and removed by some strategy, the performance of the semi-supervised algorithms should be improved. However, such useful techniques of identifying and removing noise have been seldom explored in existing semi-supervised algorithms. In this paper, we use the semi-supervised ensemble method "Co-forest" with data editing (we call it CEWS-Co-forest) to improve sparsely labeled medical dataset. The cut edges weight statistic data editing technique is used to actively identify possibly mislabeled examples in the newly-labeled data throughout the co-labeling iterations in Co-forest. The fusion of semi-supervised ensemble method with data editing makes CEWS-co-Forest more robust to the sparsity and the distribution bias of the training data. It further simplifies the design of semi-supervised learning which makes CEWS-co-forest more efficient. An experimental study on several medical data sets shows encouraging results compared with state-of-the-art methods.

Keywords-semi supervised learning, data editing, Co-forest, Ensemble methods, medical diagnosis.

1 Introduction

The goal of semi-supervised learning lies in understanding the combination of labeled and unlabeled data; this can alter learning behavior and design algorithms that benefit from such a combination. Semi-supervised learning is of great interest for automatic learning and data mining because it can use unlabeled data that are readily available on stage to improve supervised learning tasks when labeled data are scarce or expensive to obtain.

The study of semi-supervised learning is motivated by two factors: its practical value in building better computer algorithms, and its theoretical value in understanding learning in machines and humans. Semi-supervised learning has considerable practical value. In many tasks, there is a shortage of labeled data. Labels can be difficult to obtain because they require human experts, special devices and slow expensive experiments.

DOI: 10.14738/tmlai.54.3290
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3290

Most semi-supervised learning strategies are based on the extension of supervised or unsupervised learning to include additional information typical of the other learning regime. More specifically, semi-supervised learning encompasses several different contexts, including:

- Semi-supervised classification: The purpose of the semi-supervised classification is to teach the hypothesis h from labeled and unlabeled data; so that it is preferable to the used hypothesis to be supervised and trained on the label data.
- Semi-Supervised Clustering: The objective here is to adapt existing clustering methods to support constraints or labeled data, in order to produce clusters for unlabeled data using the supervised information.
- The Active learning: refers to methods that select unlabeled examples that are the most important, and an oracle can be proposed for the labeling of these instances; the objective is to minimize the labeling data (Freund, Y. et al., 1997). Sometimes it is called selective sampling or sample selection [20].

Semi-supervised learning is the considered solution to the problem of manual annotation difficulty because it can use both labeled and unlabeled data to give us a more reliable estimate of the decision limit. Intuitively, the distribution of untagged data helps to identify regions with the same label, and the few labeled data provide then the actual labels. From a different perspective, semi-supervised learning can achieve the same level of performance as supervised learning, but with fewer labeled examples. This reduces the annotation effort, which leads to a reduced cost.

Many semi-supervised ensemble learning (SSL) algorithms have been proposed, among which the ”Co-forest” algorithms are widely used. In this work, we present an improvement of Co-forest algorithm. It uses a filtering method to identify and correct the examples possibly mislabeled throughout the co-labeling iterations. Despite all the improvement of Co-forest by data editing techniques, our method CEWS Co-forest proposes a filtering method that permits to remove mislabeled examples through the control of neighborhood data, where the local cut edge weight statistic is used to help estimate whether a newly labeled example is reliable or not, and only the reliable examples are used to enlarge the labeled training set. We have conducted a comparative study that has indicated an overall significant improvement of our method compared to the existing data editing semi-supervised approach.

This paper isorganized as follow: a review of some ensemble methods in the semi-supervised field is performed in section2. We then describe in section 3, the Co-forest algorithms and its existing data-editing version DE-Co-forest. Section 4 is about the general process of our proposed approach by local cut edge weight statistic editing training data. We validate our algorithm and the choices we have made in an experimental phase in section 5. Finally, we endup with a conclusion that summarizes the contributions made and the tracks defining possible opportunities for future work.

2 Ensemble Methods in Semi-Supervised Learning

The first to be born in this category is the co-Training algorithm proposed by Blum and Mitchell [4] for semi-supervised classification web pages.

The idea of co-learning is that the feature space can be divided into two sub-spaces providing each a good learning environment. Thus, initially two classifiers are trained with the labeled data on two different subspaces. Then, each obtained classifier for each subspace is used to determine the probable class of the unlabeled data that will be used to re-train the other classifier.

To use the co-training method, you must have two different views of data to be classified, and these two different views have to be compatible and independent, and each view is used to extract the characteristics, so each view gives rise to a characterization Different from the forms to be recognized.

Compatibility makes it possible to have the same label for a given example according to each view considered independently. As far as independence is concerned, one wants for a given example, that there is no correlation between the characteristics resulting from the two different views. However, in practice, it is not always possible to obtain two independent subsets of attributes relative to the label, which makes it difficult to generalize co-Training.

To overcome this difficulty, Li and Zhou introduced in [7] a new algorithm that extends the paradigm of co-Training using Random Forest [6]. This algorithm named Co-forest uses N classifiers. N−1 classifiers are used to determine examples of trust, called concomitant Set Hi = HN-1. The confidence measure of the newly labeled example can be simply estimated by the degree of agreement on labeling, i.e. the number of classifiers that agree on the label assigned by Hi.

The approaches proposed by Blum and Mitchell [4], Zhou and Li [9] show the advantage of using multiple classifiers. Which means, learning these classifiers involves predicting the unlabeled examples before using them. Therefore, the algorithm of Li and Zhou, offers the best compromise in the semi supervised approach. Jiang and Zhou in their paper [11], provide an improvement Co-forest with very interesting results by integrating an filtering method ”DATA Editing”, the algorithm is calledDE-Co-forest.

DE-Co-forest uses the RemoveOnly editing approach [11] to identify and eliminate the ”suspect” noisy mislabeled examples in the subset of newly certified learning ones.

The data editing (filtering) approaches have the advantages that they are very efficient and robust against the over fitting. However, they tend to select examples with rather than redundant information, and do not take into account the interactions between the elements.

Therefore, in this work we are also interested in improving Co-forest, but unlike the RemoveOnly editing approach DE-Co-forest, which is based on calculating distance to the removal of noisy elements, we propose a filtering method that permits to remove mislabeled examples through the control of neighborhood data, with the local cut edge weight statistic graph strategy.

3 Methods

Many semi-supervised learning (SSL) algorithms have been proposed, among which the ”Co-forest” algorithms are widely used.

3.1 Co-forest algorithm

Co-forest was proposed by Li and Zhou [7]. This proposal is an extended version of the co-Training paradigm [4] by the ensemble method Random Forest [6]. In Co-forest, a set of N classifiers is used instead of two in co-Training. In this way, we can effectively improve the confidence estimate by each classifier. If we want to consider the labeled instance, the most confident by a classifier hi (i = 1, 2 . . . N) of the set H*, the all other classifiers are used except hi, called concomitant ensemble of hi and denoted by Hi. Therefore, the confidence level is calculated as the degree of agreement on the label, i.e. the number of classifiers agree on the label assigned by Hi. The general idea of Co-forest is to learn a set of classifiers.

More specifically, Co-forest is an iterative process, the concomitant ensemble hi will test each unlabeled example. Thereafter, if the number of classifiers that agree on a particular label exceeds a predefined threshold θ, this new label is assigned to the example and then it will be copied in the new labeled set. In the next iteration, the new labeled set is used for refining hi. Hereafter, we note that the unlabeled examples are not deleted, so they can be selected by other Hj (j≠i) in the following iterations. Consult [7, 20] for more details on the understanding of co-Forest.

3.2 DE-Co-forest algorithm

In semi-supervised learning, there is a problem that may affect Co-forest as well as other algorithms such as co-Training, which is the unlabeled examples may be mislabeled and introduced into the learning process. This is due to the limited number of examples initially labeled that usually generates low classifiers, which lacks precision and diversity. Based on this observation, a new algorithm that combines Co-forest with a data editing technique called DE-Co-forest is used. DE-Co-forest uses a data editing technique to identify and possibly eliminate mislabeled examples through iterations of co-labeling. In DE-Co-forest the RemoveOnly data editing technical [11] is used to identify mislabeled data.

Its principle is that the label of each unlabeled instance is not only determined by multiple classifiers, but also by the nearest neighbor rule. If the label is compatible with those selected by a minimum of k' nearest neighbor data, the unlabeled instance data with the greatest confidence are added to the training set. Otherwise, they are rejected and removed from the set of re-learning.

This method drove improvements to enrich the learning set, which is based on the k-nearest neighbor (k-NN). In this context, Cover and Hart [12] studied the asymptotic optimality of the nearest neighbor (NN) rule [13] and they proved that the NN rule is asymptotically optimal when different classes do not overlap in the input space. Otherwise, it may seem as one of the sub-optimality of the NN rule and it can overcome a bad classification. To decrease this error of the optimality, we propose the implication of CEWS [14] filtering method that permits to remove mislabeled examples through the control of neighborhood data, where the local cut edge weight statistic is used to help estimate whether a newly labeled example is reliable or not. Thereby, only the reliable examples are used to enlarge the labeled training set. We conduct a comparativestudy that indicate an overall significant improvement of our method compared to the existing data editing semi-supervised approach.

4 The Proposed Method CEWS Co-Forest Algorithm

The Cut edges weight statistic (CEWS) [14], this filtering method permits to remove mislabeled examples through the control of neighborhood data. At the beginning, it is necessary to build a geometrical connected graph like Toussaint's Relative Neighborhood Graph [15] on all examples of the training set. By definition a neighborhood graph G = (V, E) [16] is represented by vertex V and there exists an edge E between two vertices's xi and xj if the distance between xi and xj satisfies Eq. (1).

$$R_i = J_i / I_i \tag{2}$$

Where, Ii is the sum of weights relative to edges for sample xi Eq(3), Ji is the sum of weights relative to cut edges for sample xi Eq(4) and wij is the weighting distance of each edge Eq(5).

$$I_i = \sum_{j \in Neighborhood(x_i)} w_{ij} \tag{3}$$

$$J_i = \sum_{j \in Neighborhood(x_i), y_j \neq y_i} w_{ij} \tag{4}$$

$$w_{ij} = 1/\left(1 + dist\left(x_i, x_j\right)\right) \quad (5)$$

5 Experiments and Results

We have selected a set of seven databases from ASU repository [18] and UCI [19] their characteristics are summarized in Table I. To study the effectiveness of CEWSCo-forest incomparison to the performance of Co-forest and DE-Co-forest.

Table 1. The average accuracy of the compared algorithms under different labeled rateμ

Labeled rate	*Methods*	*data_C*	*Leukemia*	*Lung*	*Musk*	*Ovarian*	*Pancreatic*	*Prostate*
μ=80%	**Co-Forest**	0,5800	0,8615	0,8732	0,8138	0,8111	0,5452	0,8457
	DE-Co-Forest	0,5500	0,8538	0,8648	0,8175	0,8444	0,5194	0,8800
	CEWS-Co-Forest	0,5900	0,9077	0,8761	0,8200	0,8889	0,5097	0,9029
μ=60%	**Co-Forest**	0,6100	0,8615	0,8507	0,8213	0,8000	0,4871	0,8800
	DE-Co-Forest	0,6100	0,8308	0,8507	0,8088	0,8222	0,4968	0,8400
	CEWS-Co-Forest	0,6700	0,8846	0,8648	0,8250	0,8333	0,4903	0,8571
μ=40%	**Co-Forest**	0,5800	0,8154	0,8225	0,7763	0,8444	0,5032	0,7600
	DE-Co-Forest	0,5800	0,8154	0,8338	0,7788	0,8444	0,5032	0,7200
	CEWS-Co-Forest	0,5900	0,8538	0,8394	0,7875	0,8667	0,5194	0,7714
μ=20%	**Co-Forest**	0,6400	0,7615	0,7352	0,7138	0,5333	0,4903	0,6800
	DE-Co-Forest	0,6500	0,7538	0,7493	0,6988	0,5556	0,4903	0,6686
	CEWS-Co-Forest	0,6800	0,8154	0,7718	0,7188	0,5778	0,5194	0,6857

Table 2. Description of Experimental High Dimensional Datasets

Datasets	*#instances*	*#features*	*#class*
Data C	60	7130	2
Leukemia	73	7129	2
Lung	203	12600	5
Musk	476	166	2
Ovarian	54	1536	2
Pancreatic	119	6771	2
Prostate	102	12533	2

For each dataset, a 10 cross validation is carried out for evaluation. The training data are randomly divided into two sets: L labeled and unlabeled U determined by a rate (μ), which is calculated by the size of L on the size of L U U. To simulate different amounts of unlabeled data, four different unlabeled rates μ = 20 %, 40 %, 60 % and 80 %, are studied.

The distributions of class in L and U are maintained similar to the original set. In these experiments, the value of N is 6 trees. Confidence level θ is set at 0.75, i.e., a newly labeled example is considered trusted if more than three quarters of the trees are agreements on its assigned label. For the RemoveOnlydata editing method, we fixed the number of neighbors k at 3, and the minimum number of neighbors equals to 2.

In term to estimate the accuracy on each dataset, we have predetermined a set of labeled examples. For each set, the algorithm is evaluated on its ability to correctly predict the labels of unlabeled examples. The labeled samples were randomly selected, with the only constraint being the presence of at least one example of each class for each set.

To compare the performances of CEWS Co-forest to Co-forest and DE-Co-forest, for each dataset with a specific labeled rate μ, a cross-validation is repeated ten times, and the results are averaged and recorded. Table 2 shows the average accuracy results and the ranking of each testing algorithm obtained. Specifically, it shows the overall results of the analyzed algorithms over the seven used datasets with 20, 40, 60 and 80 % rate of unlabeled data.

Our proposition outperforms the other algorithms; except CEWS-Co-Forest's performance was degraded on Pancreatic dataset with 80% and 60% of unlabeled rate (Table II). This can be explained by poor learning of the initial hypothesis and by the addition of misclassified data especially in training set. However, the CEWS-Co-Forest benefits much from the unlabeled data since the performances are evidently improved over all the seven datasets compared to other algorithms.

The immersion of the Cut edges weight statistic (CEWS) into the co-forest algorithm process allows enhanced confidence labeling to improve the classification accuracy, so we can deduce that CEWS Co-forest algorithm gives a good result in comparison with other algorithms.

6 Conclusion

The presented algorithm is an improvement of the Co- forest method [7] for semi-supervised classification. The aim of data editing in CEWS Co-forest is to identify and remove the noise contained in labeling step and thus to improve the overall performance. Our basic consideration is to implement a filtering method that permits to remove mislabeled examples through the control of neighborhood data, while, we are fully utilizing the advantage of ensemble learning in order to incur less computation complexity when improving the accuracy.

Experiments on high biomedical data sets show that data editing is a very useful technique for improving the performance of sparsely labeled data classification, and it makes the algorithm more efficient. For future work, we will further explore new techniques to cope with the training data sparsity and trainingdata bias for sparsely labeled data classification, e.g. semi-supervised clustering aided techniques.

REFERENCES

[1] O. Chapelle, B. Sch¨ and A. Zien, Semi-Supervised Learning,MIT Press, Cambridge, MA, 2006.

[2] Antoine Cornu´ and Laurent Miclet, Apprentissage artificiel : Concepts et algorithmes, Eyrolles, June 2010.

[3] Xiaojin Zhu, "Semi-Supervised learning literature survey," Tech. Rep.,Computer Sciences, University of Wisconsin-Madison, 2005.

[4] Avrim Blum and Tom Mitchell, "Combining labeled and unlabeled datawith co-training," in Proceedings of the eleventh annual conference onComputational learning theory, New York, NY, USA, 1998, COLT' 98,pp. 92–100.

[5] L. G. Valiant, "A theory of the learnable," Commun. ACM, vol. 27, no.11, pp. 1134–1142, Nov. 1984.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5–32,2001.

[7] Ming Li and Zhi-Hua Zhou, "Improve computer-aided diagnosis withmachine learning techniques using undiagnosed samples," Trans. Sys.Man Cyber. Part A, vol. 37, no. 6, pp. 1088–1098, Nov. 2007.

[8] Yan Zhou and Sally Goldman, "Democratic co-learning," in Proceedingsof the 16th IEEE International Conference on Tools with ArtificialIntelligence, Washington, DC, USA, 2004, ICTAI '04, pp. 594–202,IEEE Computer Society.

[9] Zhi-Hua Zhou and Ming Li, "Tri-training: Exploiting unlabeled datausing three classifiers," IEEE Trans. on Knowl. and Data Eng., vol. 17,no. 11, pp. 1529–1541, Nov. 2005.

[10] Boaz Leskes and Leen Torenvliet, "The value of agreement a newboosting algorithm," J. Comput. Syst. Sci., vol. 74, no. 4, pp. 557–586,June 2008.

[11] Yuan Jiang and Zhi hua Zhou, "Editing training data for knn classifierswith neural network ensemble," in Lecture Notes in Computer Science,Vol.3173. 2004, pp. 356–361, Springer.

[12] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEETrans. Inf. Theor., vol. 13, no. 1, pp. 21–27, Sept. 2006.

[13] Fix and J. L. Hodges, "Discriminatory analysis. Nonparametricdiscrimination: consistency properties," Jr. U.S. Air Force Sch. AviationMedicine, Randolf Field, vol. Rep.4, pp. Project 21–49–004, ContractAF 41 (128)–31, 1951.

[14] Fabrice Muhlenbach, St[U+FFFD]ane Lallich, and Djamel A. Zighed,"Identifying and handling mislabelled instances.," J. Intell. Inf. Syst.,vol. 22, no. 1, pp. 89–109, 2004.

[15] Godfried T. Toussaint, "The relative neighbourhood graph of a finiteplanar set," Pattern Recognition, vol. 12, pp. 261–268, 1980.

[16] L. Devroye, L. Gy and G. Lugosi, A Probabilistic Theory of PatternRecognition, Springer, 1996.

[17] Yu Wang, Xiaoyan Xu, Haifeng Zhao, and Zhongsheng Hua, "Semi-supervised learning based on nearest neighbor rule and cut edges,"Know.-Based Syst., vol. 23, no. 6, pp. 547–554, Aug. 2010.

[18] Reza Zafarani and Huan Liu, "Asu repository of social computingdatabases," 1998.

[19] D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, "Uci repositoryof machine learning databases," 1998.

[20] Lytras, Miltiadis D., and Paraskevi Papadopoulou. "Applying Big Data Analytics in Bioinformatics and Medicine." IGI Global, 2018. 1-402. Web. 6 Apr. 2017. doi:10.4018/978-1-5225-2607-0.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Psychological and Physiological Profiles in Oncology Caregivers: a Multivariable Cross-Sectional Study

Esther García[a], Sira Aguiló[b], Adriana Arza[a,c], Jorge M. Garzón-Reya[c], Jordi Aguiló[a,c]
[a]CIBER-BBN, Barcelona, Spain, [b]Hospital Clínic, Barcelona, Spain, [c]Universitat Autònoma de Barcelona, Barcelona, Spain
acmlis.conference@gmail.com

ABSTRACT

Scientific research has led to enhanced medical assistance, and hence higher disease control and a continuous lifespan dilation. However, research mainly focusses on the patient's condition, and, in some cases, caregivers' health is undervalued. In this paper, a close-related emotionally involved caregiver group is examined and compared to a non-caregiver group in order to investigate the role of chronic stress in their psychological and physiological state. For this purpose, blood, saliva, and hair samples, physiological signals and three validated psychometric tests are collected from both groups. Results from 39 subjects show interesting significant differences among all the psychometric profiles between the two groups although no sufficient biochemical data supports the relation with potential health consequences.

Keywords: Oncology caregivers; Stress biomarkers; Chronic stress; Psicometric tests

1 Introduction

Scientific research has led to enhanced medical assistance, and hence higher disease control and a continuous lifespan dilation. However, research mainly focusses on the patient's condition, and, in some cases, caregivers' health is undervalued. Informal caregivers are non-remunerated and non-professional people in charge of the care of a family member with limited autonomy due to physical or cognitive impairment.

Daily caregiving leads to a burden escalation with physical and psychological impact on the caregiver health [1,2]. The unrelenting repercussions on caregivers' health come from the physical burden related to the daily tasks, the psychological frustration after the inevitable deterioration of the cherished patient and even the funding of the care expenses can passively impact a person's wellbeing [3,4] Common symptoms are sleeping disturbance, irritability, fatigue, weight gain or loss without diet, frequent colds and infections among others [3,5]. From a psychological point of view, the subjective reaction towards the care receptor demands often lead to psychological problems such as depression, anxiety and stress, collectively referred to as the caregiver stress syndrome [3,5,6]. Some studies have shown that stress is an independent risk factor of acute myocardial infarction, comparable to arterial hypertension and abdominal obesity, diabetes mellitus type 2, psychiatric diseases and implication in the development of Alzheimer's disease [7-10].

DOI: 10.14738/tmlai.54.3291
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3291

Stress factors induce homeostasis disturbance in the organism and the coping mechanisms to recuperate the equilibrium are known to be mainly regulated through Autonomic Nervous System and the Hypothalamus-Pituitary-Adrenal axis (HPAA) [11]. These adaptations include endocrine, electrophysiological and/or psychological changes that become pathologic if they are not limited to an adaptive and short-in-time response.

The fastest response noticeable are the electrophysiological signs coming from the activation of the autonomic nervous system. The information of the balance between the activation of the sympathetic and parasympathetic nervous systems can be provided by a wide variety of measures; for instance, the heart rate and its components: heart rate variability (HRV) low and high frequency ratios (LF/HF) and R to R peaks variability (RRV) [12-14]. There are also reports of an increase in the respiratory rate (RR) and its variability (RRV) in front of a stressor [15]. Other studies have established a relationship between stress and blood pressure (BP), the sweat-dependent electrical characteristics of the skin (GSR – galvanic skin response) and the skin temperature (ST) [16-19].

A parallel slower response is mediated through the HPAA initiated in the paraventricular nucleus of the hypothalamus. This endocrine response can be revealed by means of several biomarkers. For instance, reported data show quick variations in blood glucocorticoid levels (cortisol in humans) due to the action of stressors [20]. However, salivary and hair samples seem to be growing non-invasive and more stable measures of cortisol concentrations [21,22]. Similarly, α-amylase is known to be a marker of the activity of the sympathetic nervous system in response to stress. Copeptin, a surrogate biomarker of vasopressin (AVP), is another possible marker of stress, and it is quite stable both in circulation and ex-vivo. Levels of copeptin are sensitive to situations involving physical stress [23], but could also be sensitive to emotional stress [24]. Inflammation biomarkers such as interleukin 6 and 10 (IL6, IL10) and tumor necrosis factor alpha (TNF-α) have been associated with depressive disorders as well [25].

Considering the psychological branch, psychometric tests are the most commonly accepted tools for measuring emotional distress. The PSS, STAI and VASS are validated, endorsed and very well documented by the medical world [26-28]. They are the reference standard for helping to determine the level of stress reached by the subject in each situation.

In spite of the wide variety of studies looking for a unique parameter to measure stress, there are several throwbacks. For instance, the multidimensional nature of the questioners and the lack of standardization hinder the identification of the most appropriate test to measure objectively the stress level [29]. Additionally, since the HPA axis is a highly adaptive system which is characterized by marked inter- and intraindividual variability, the development of such markers of HPA axis regulation in humans is still a rather challenging task.

Bearing all this information in mind, and considering that nowadays, there is still not an objective measure of signs that allow the early diagnosis of stress, ES3 project aims to find a multivariable approach, quantitatively and objectively assessing the three branches affected, that could provide very valuable information to detect and prevent symptoms and diseases associated with the acute and chronic stress [30].

The present study shows some preliminary results of a pilot included in the bigger project mentioned above [30]. This pilot is focused in chronic stress in oncology caregivers, evaluating biochemical, electrophysiological and psychometric data. According to the World Health Organization, cancer is the second leading cause of death globally, and for many oncology patients, palliative care is the only available treatment usually involving highly demanding home-based care. Early diagnosis of the caregiver's stress syndrome along with home care programs and nursing support could prevent chronic stress effects and reduce the development of stress-related diseases in this population.

The aim of this work is to assess the perceived stress and endocrine, metabolic and immunologic profile comparing an oncology informal caregivers group with a matched comparison group.

2 Patients and Methods

2.1 Patients

Caregivers of oncologic patients (oncology caregivers) attending at the Oncology Service of the Hospital Clínic of Barcelona (HCPB, Barcelona, Spain), were invited to participate in the cross-sectional study. A paired group of non-caregivers volunteers unrelated with the patients were invited to participate in this study.

The study was performed in the Hospital Clínic of Barcelona between September 2015 and June 2016. The study was approved by the institutional Committee of Ethics and in accordance with those of the World Medical Association and the Helsinki Declaration. All participants were informed about the purpose and characteristics of the study and their rights to withdraw of the study at any time during process. All participants signed the informed consent.

Inclusion criteria for caregivers were: > 18 years old, to take care of an oncological patient with progressive incurable cancer and oncospecific treatment, to be unpaid for the caregiving tasks, to live with the patient in the same household, to be emotionally involved (family or partner), and to develop the caregiver activity for a minimum period of 6 months.

Inclusion criteria for non-caregivers were: to be at least 18 years old and not to be in charge of any family requiring care. Non-caregivers were matched for age (up to 5 years older or younger), gender and body mass index (BMI).

Exclusion criteria for all participants were any of: failure to meet any of the inclusion criteria; habitual intake of psychotropic or corticoid; to be diagnosed with alteration of the hypothalamic axis adrenal pituitary; or to have a Body Mass Index (BMI) greater than 35.

2.2 Endpoints and Methods

Socio-demographic and lifestyle variables analysed were: age, gender, BMI, habits (smoking status, coffee and alcohol intake, consume of drugs and physical activity), health condition (chronic disease diagnosed, medication and psychotropic treatment) and other variables (cohabitation condition and relationship and job status). Detailed characteristics of these variables are shown in Table 1.

Primary endpoints were perceived stress scores using Spanish version of psychometric tests: Perceived Stress Scale (PSS) [26], State-Trait Anxiety Inventory (STAI) [28], and Stress Visual Analogue Scale (VASS) [27].

PSS provides a global measure of perceived stress in a situation appraised as stressful and considering the feelings and thoughts occurred during the last month. The test consists in a 10 questions survey with liker scaling responses (never, almost never, sometimes, fairly often and very often). STAI measures a subjective emotional state. The test consists of 40 total questions with liker scaling responses, 20 out of 40 total questions measure the trait anxiety (almost never, sometimes, often and almost always) and the remaining 20 out of 40 total questions measure the state anxiety (not at all, somewhat, moderately so, very much so). VASS consists in a simple visual 100-point scale (0, not at all; 100, absolutely stressed), in which participants rate their present perceived stress. Tests take few minutes for the patient to complete and no special education or training is required to administer them.

Secondary endpoints were endocrine, metabolic and inflammatory markers: plasma copeptin (pmol/L), copeptin osmolarity (mOsm), glucose (mg/dL), glycated hemoglobin (HbA1c; IFCC and NGSP/DCCT units) (%), LDL (mg/dL), HDL (mg/dL), cholesterol (mg/dL), triglycerides (mg/dL), salivary α-amylase (U/mL), salivary cortisol (µg/dL), hair-cortisol (pg/mg), TNFα (pg/mL), IL-6 (pg/mL) and IL-10 (pg/mL). These analytes can result in stress biomarkers and information of the overall health of the subjects.

Plasma samples were obtained from 10 mL of blood collected in serum tubes for the biochemical and immunological analysis and ethylenediaminetetraacetic acid (EDTA) tubes for the HbA1c and copeptin analysis. Salivary samples were collected with salivettes (Sarstedt, Granollers, Spain) and stored at -20ºC until the analysis day. Then, salivettes were thawed at room temperature and centrifuged at 4ºC for 10 minutes at 1000g. Salivary α-amylase was analysed using Salimetrics Salivary Alpha-Amylase Assay (Salimetrics®, Carlsbad, CA, USA), per duplicate using the same kit batch and following the instructions of the manufacturer. The amount of α-amylase activity is directly proportional to the 405nm absorbance. Salivary cortisol was analysed using Salimetrics Cortisol Enzyme Immunoassay (Salimetrics®, Carlsbad, CA, USA), by duplicate using the same corresponding kit batch and following the instructions of the manufacturer. The limit of quantification of cortisol assay is 0.012 µg/dL.

Hair samples were stored at room temperature in propylene tubes protected from light. The accumulated cortisol was extracted according to the procedure by Scorrano et al. with minor modifications [31]. Briefly, the 3 cm of hair closest to the scalp were obtained and washed two times for 1 minute in 10 ml of isopropanol and completely air-dried at room temperature. For the extraction of cortisol, 40 mg of hair was mixed with 1.6 mL of methanol overnight by continuous rotation (20 rpm/min). After centrifugation (10,000g, 10 min, 4°C), the methanol was recovered in a new clean glass tube and the procedure was repeated again. Total recovered methanol was pooled and dried up under a nitrogen stream. Cortisol extracts were reconstituted in 200 µL of sodium phosphate buffer and analysed by using Salimetrics® Cortisol Enzyme Immunoassay. The template is used to format your paper and style the text. All not revise any of the current designations.

All the biochemical analysis were processed at the Biomedical Diagnosis Centre at the Hospital Clínic of Barcelona. Cortisol and α-amylase in saliva and hair samples were processed in the laboratory of the Endocrinology and Radioimmunoanalysis Service of the Animal Physiology Unit at the Universitat Autònoma de Barcelona, and TNFα, IL-6 and IL-10 were sent to the Myeloid Cell Laboratory at the Biological Research Centre in Madrid.

2.3 Data analysis

Descriptive analysis of data was performed with SAS® software version 9.4 and Matlab R2014b. For categorical variables, percentages of events were calculated and the comparisons of groups were analysed by Fisher's exact test. For quantitative variables, the median and the inter-quartile range (IQR) were calculated. Comparisons among groups were analysed by non-parametric analysis (Wilcoxon Signed Rank Test) according to the sample size and criteria of normality (Shapiro-Wilk test). All tests were two-tailed, and p-values < 0.05 were considered statistically significant.

3 Results

A total of 39 subjects participated in the study (19 non-caregivers and 20 oncology caregivers). The median age (IQR) of participants was 55 years (45-64), the median (IQR) of BMI was 25.56 (21.60- 28.39) and there was a majority gender group with 66.67% of women. Non-significant differences were found among groups in terms of gender, age and BMI (Fisher's Exact and Wilcoxon signed rank tests p-value >0.05) (Table I).

Table I. Sociodemographic and lifestyle characteristics.

	Non-Caregivers	*Oncologic Caregivers*	*Total*
Participants	19 (48.72)	20 (51.28)	39 (100.00)
Age. yr	53.00 (45.00- 65.00)	55.50 (47.00- 62.00)	55.00 (45.00- 64.00)
Sex			
Men	6 (31.58)	7 (35.00)	13 (33.33)
Women	13 (68.42)	13 (65.00)	26 (66.67)
BMI	25.61 (20.83- 28.39)	25.40 (22.24- 28.04)	25.56 (21.60- 28.39)
Chronic disease			
Non	15 (78.95)	8 (40.00)*	23 (58.97)
Yes	4 (21.05)	12 (60.00)*	16 (41.03)
Medication			
Non	11 (57.89)	9 (45.00)	20 (51.28)
Yes	8 (42.11)	11 (55.00)	19 (48.72)
Psychotropic			
Non	17 (89.47)	14 (70.00)	31 (79.49)
Occasionally	2 (10.53)	6 (30.00)	8 (20.51)
Smoker			
Non	18 (94.74)	12 (60.00)*	30 (76.92)
Yes	1 (5.26)	8 (40.00)*	9 (23.08)
Daily Coffee			
None	1 (5.26)	1 (5.00)	2 (5.13)
≥ 1 cup	18 (94.74)	19 (95.00)	37 (94.87)
Daily Alcohol			
None	2 (10.53)	9 (45.00)*	11 (28.21)
≥ 1 glass	17 (89.47)	11 (55.00)*	28 (71.79)
Sport practice			
Non	3 (15.79)	15 (75.00)	18 (46.15)
(2) Yes	16 (84.21)	5 (25.00)	21 (53.85)

Married			
Non	8 (42.11)	5 (25.00)	13 (33.33)
Yes	11 (57.89)	15 (75.00)	26 (66.67)
Cohabitation			
Family	6 (31.58)	8 (40.00)	14 (35.90)
Partner	8 (42.11)	9 (45.00)	17 (43.59)
Alone	5 (26.32)	3 (15.00)	8 (20.51)
Job status			
Inactive	4 (21.05)	10 (50.00)	14 (35.90)
Active	15 (78.95)	10 (50.00)	25 (64.10)

[1] Less than twice/week and any during the last 24hrs. [2]Either occasionally or currently. Abbreviations: yr, years. (*) Group comparison statistically significant (p-value <0.05). Fisher's exact test for discrete variables, N(%) and Wilcoxon Signed Rank Test for quantitative variables, Median (IQR).

Even though the two groups showed similar medication and psychotropic consumption (Fisher's Exact Test, p= 0.1842 and p=0.2351, respectively), the oncology caregivers' group manifested a higher chronic disease incidence of 60% compared to the 21.05% of the non-caregiver group (Fisher's Exact Test, p= 0.0225). Both groups presented similar social relations, with similar cohabitation, relationship and job status (Fisher's Exact Test, p-values > 0.05). More than 1 cup of coffee per day was the usual intake of the 95% (37/39) of total participants (Fisher's Exact Test, p= 1.0000). A 40% (8/12) of caregiver subjects were smokers, a significantly greater proportion than in the comparison group where there was only 1 out of 19 (5.26%, Fisher's Exact Test, p= 0.0197). On the contrary, the highest percentage in daily alcohol intake was in the non-caregivers group (89.47%, Fisher's Exact Test, p=0.0310). A 53.85% (21/39) of total participants practiced sport. However, the percentage of physically active subjects was considerably bigger in the comparison group than in the caregiver's group (84.21% and 25% respectively, Fisher's Exact Test, p=0.0003) (Table I).

Table 2. Psychometric tests scores

	Non-caregivers ***Median (IQR)*** N = 19	***Oncologic caregivers*** ***Median (IQR)*** N = 20	***Wilcoxon Signed Rank Test***
PSS	22.00 (18.00-23.00)	24.50 (23.00-27.50)	S = 75.5; p=0.0012*
STAI-s	11.00 (6.00-16.00)	47.50 (35.00-53.00)	S = 95; p<0.0001*
STAI-t	13.00 (9.00-21.00)	28.50 (20.00-40.00)	S = 80.5; p=0.0004*
VASS	40.00 (23.00-50.00)	80.00 (65.00-82.50)	S = 95; p<0.0001*
[2]Summary score	99.00 (67.00-116.00)	218.50 (190.00-249.00)	S = 94; p<0.0001*

[1] Summary score obtained from the sum of psychometric test scores of each participant for a general overall analysis. [*] Group comparison statistically significant (p-value <0.001).

Perceived stress scores analysed showed significantly different self-perceived stress profiles in all the psychometric tests given. We found evidences that the level of perceived stress in oncology caregivers is higher than the non-caregivers (Wilcoxon signed rank tests p-value <0.001). Oncology caregivers presented the highest score distribution in all psychometric tests: PSS test (median: 24.50; IQR: 23.00-27.50), STAI-s test (median: 47.50; IQR: 35.00-53.00), STAI-t test (median: 28.50; IQR: 20.00-40.00) and in VASS rate (median: 80.00; IQR: 65.00-82.50) (Table II). The differences between the scores in both groups

were the more noticeable in the STAI-s test and the VAS scale. The summary score is obtained from the sum of psychometric test scores of each participant for a general overall analysis and this combination shows the most pronounced difference (Fig. 1).

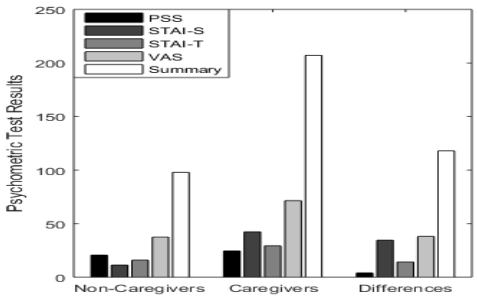


Figure 1. Psychometric test results from both the oncology caregivers and comparison groups and the score differences

Table 3. Endocrine, metabolic and inflammatory analytes concentrations.

	Non-caregivers Median (IQR)	*Caregivers Median (IQR)*	*Wilcoxon Signed Rank Test*
Copeptin (pmol/L)	5.45 (2.25-11.50)	6.20 (3.50-8.60)	S = 15.5; p=0.5226
Copeptin Osmolarity (mOsm);	300.00 (296.00-304.50)	303.00 (298.00-307.00)	S = 43; p=0.0614
Glucose (mg/dL)	93.00 (89.00-101.00)	101.00 (90.00-132.00)	S = 41.5; p=0.0976
HbA1c (IFCC) (%)	5.00 (4.90-5.20)	5.20 (4.70-5.60)	S = 20; p=0.3992
HbA1c (NGSP/DCCT) (%)	5.40 (5.30-5.60)	5.60 (5.10-6.00)	S = 20; p=0.3992
LDL (mg/dL)	109.70 (100.00-135.40)	127.60 (104.60-154.00)	S = -22.5; p=0.3465
HDL (mg/dL)	58.00 (51.00-67.00)	58.00 (53.00-71.00)	S = -31.5; p=0.1811
Cholesterol (mg/dL)	199.00 (172.00-214.00)	208.00 (181.00-244.00)	S = 8.5; p=0.7454
Triglycerides (mg/dL)	81.00 (58.00-96.00)	77.00 (57.00-156.00)	S = 15; p=0.5678
Salivary α-amylase (U/mL)	198.65 (98.15-217.25)	260.78 (135.90-382.05)	S =49; p=0.0494*
Salivary cortisol (µg/dL)	0.16 (0.12- 0.21)	0.14 (0.08-0.21)	S = 7; p=0.7983
Hair-cortisol (pg/mg)	11.52 (7.76-16.51)	9.46 (6.70-18.88)	S =-8; p=0.7680
TNFa (pg/mL)	0.73 (0.00-1.56)	0.42 (0.00-1.56)	S = 14.5; p=0.4288
IL-6 (pg/mL)	1.15 (0.00- 3.46)	2.40 (0.00-5.38)	S =11.5; p=0.5016
IL-10 (pg/mL)	3.00 (2.27-3.74)	3.15 (1.53-3.74)	S =-23.5; p=0.2787

(*) Group comparison statistically significant (p-value <0.05).

With respect to the biochemical analytes studied, the only levels that significantly differed were the α-amylase concentrations (Wilcoxon Signed Rank Test, p-value = 0.0494). Distribution values of concentration of glucose, glycated hemoglobin, copeptin, LDL, HDL, cholesterol, triglycerides, salivary

cortisol, hair-cortisol, TNFα, IL-6 and IL-10 did not differ statistically significant between non-caregivers and oncology caregivers groups (Wilcoxon Signed Rank Test, p-value >0.05) (Table III). Several clinically relevant concentrations were found in both groups although mainly in the caregivers' group (normal values of the CDB: glucose 65-110 mg/dL, HbA1c <6.5% IFCC, LDL <180 mg/dL, HDL >40 mg/dL, total cholesterol 148-247 mg/dL, triglycerides 50-150 mg/dL). We observed abnormal concentrations of glucose typed as pre-diabetes in some oncology caregivers.

4 Discussion and conclusions

The present study compares the self-perceived stress and endocrine, metabolic, and inflammatory analytes concentrations in an oncology caregiver group and a non-caregivers group. Our results provide consistent evidences that subjects with caregiver burden perceived much higher stress.

Cancer diagnosis and follow-up brings an important stress impact in cancer patients and relatives. Regular clinical visits, aggressive drug-treatments and health deterioration become more evident over time. These circumstances together with the unexpected care tasks of a family member are the root of great emotional and physical burden. PSS as well as STAI-t give a global measure of perceived stress and anxiety in front of a challenging situation considering the feelings and thoughts. The significant differences in both psychometric tests obtained reflect the tendency of this community to be more afflicted in front of situations appraised as stressful. Greater differences were observed among STAI-s and VAS scores which evaluate the stress felt at that precise moment. However, these results might be influenced by the impact of the setting for visit on oncology caregivers since they attended the study at the hospital during cancer patient visit.

Considering the analytes concentrations, our study assessed the levels of copeptin, glucose, glycated hemoglobin, LDL, HDL, cholesterol, triglycerides, α-amylase, cortisol, TNFα, IL-6 and IL-10 in oncology caregivers and non-caregivers in a unique visit. Several studies report that psychological stressors may have physiological impact in the organism [1,21,23,24]. However, almost none of the endocrine, metabolic or immunologic biomarkers selected differed significantly between the two groups in our study. Salivary α-amylase was the only analyte that showed statistically significant results. Similarly, Siddiqui A. et al. [10] found elevated levels of salivary α-amylase as well as salivary cortisol in type 2 diabetes mellitus patients as stress responses. In our study, salivary cortisol levels were within the normal range. This discrepancy of the cortisol concentrations might be biased due to the diversity of stressors, the wide repertoire of mental stress measures, the sampling strategy and the intra-individual variability of diurnal cortisol fluctuation [20,21]. Other approaches suggest hair sampling to estimate cortisol concentrations since they provide a more stable measure, but they might not exempt of cosmetic hair treatments bias [22]. Even though some studies performed in animals and humans showed that hair cortisol concentrations increase in stress situations [22,31], the results in our caregiver population disagree.

The main limitation of our study is the small size of sample enrolled in each group which limits the power of the statistical analysis and the probability of finding wider range of biomarker levels. The parameters extraction and analysis of the physiological data is currently ongoing and further results will be published.

In summary, the value of self-perceived stress is significantly greater in the caregivers' community summited to a chronic situation of stress due to the care of an oncologic patient. The work in process may provide the baseline information to initiate specific social actions addressed to increase the

empowerment of the caregivers' community. Our results ought to be consolidated in prospective cohort studies with larger groups of caregivers to help to perform medical actions directed to prevent comorbidities and social actions.

ACKNOWLEDGMENTS

Research supported by MINECO (FIS-PI12/00514 and TIN2014-53567-R) and by the Centro de Investigación Biomédica en Red sobre Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN) at the Instituto de Salud Carlos III de España.

REFERENCES

[1] Schulz R, Sherwood PR. Physical and mental health effects of family caregiving. Am J Nurs. 2008;108(9 Suppl):23-7; quiz 27. doi: 10.1097/01.NAJ.0000336406.45248.4c.

[2] Vitaliano PP, Zhang J, Scanlan JM. Is caregiving hazardous to one's physical health? A meta-analysis. Psychol Bull. 2003;129:946-72.

[3] Pinquart M, Sörensen S. Associations of stressors and uplifts of caregiving with caregiver burden and depressive mood: a meta-analysis. J Gerontol B Psychol Sci Soc Sci. 2003;58:P112-28.

[4] Oliva-Moreno J, Trapero-Bertran M, Peña-Longobardo LM, Del Pozo-Rubio R. The Valuation of Informal Care in Cost-of-Illness Studies: A Systematic Review. Pharmacoeconomics. 2017;35:331-345.

[5] Stenberg U, Ruland CM, Miaskowski C. Review of the literature on the effects of caring for a patient with cancer. Psychooncology. 2010;19:1013-25.

[6] Carmeli E. The invisibles: unpaid caregivers of the elderly. Front Public Health. 2014;2:91. doi: 10.3389/fpubh.2014.00091.

[7] Rosengren A, Hawken S, Ounpuu S, Sliwa K, Zubaid M, Almahmeed WA, et al. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364:953-62.

[8] von Känel R, Mausbach BT, Dimsdale JE, Mills PJ, Patterson TL, Ancoli-Israel S, et al. Cardiometabolic effects in caregivers of nursing home placement and death of their spouse with Alzheimer's disease. J Am Geriatr Soc. 2011;59:2037-44.

[9] Roepke SK, Allison M, Von Känel R, Mausbach BT, Chattillion EA, Harmell AL, et al. Relationship between chronic stress and carotid intima-media thickness (IMT) in elderly Alzheimer's disease caregivers. Stress. 2012;15:121-9.

[10] Siddiqui A, Madhu SV, Sharma SB, Desai NG. Endocrine stress responses and risk of type 2 diabetes mellitus. Stress. 2015;18:498-506.

[11] Selye H. Stress and the general adaptation syndrome. Br Med J. 1950; 1: 1383-92.

[12] Choi J, Ahmed B, Gutiérrez-Osuna R. Development and evaluation of an ambulatory stress monitor based on wearable sensors. IEEE Trans Inf Technol Biomed 2012; 16: 279-86.

[13] Thayer JF, Åhs F, Fredrikson M, Sollers JJ, Wager TD. A metaanalysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012; 36: 747-56.

[14] Visnovcova Z, Mestanik M, Javorka M, Mokra D, Gala M, Jurko A, et al. Complexity and time asymmetry of heart rate variability are altered in acute mental stress. Physiol Meas 2014; 35: 1319-34.

[15] Vlemincx E, Taelman J, De Peuter S, Van Diest I, Van den Bergh O. Sigh rate and respiratory variability during mental load and sustained attention. Psychophysiology 2011; 48: 117-20.

[16] Allen AP, Kennedy PJ, Cryan JF, Dinan TG, Clarke G. Biological and psychological markers of stress in humans: focus on the Trier Social Stress Test. Neurosci Biobehav Rev 2014; 38: 94-124.

[17] De Santos-Sierra A, Ávila CS, Casanova JG, Del Pozo GB. Real-time stress detection by means of physiological signals. Recent Application in Biometrics 2011; 58: 4857-65.

[18] Haufe S, Kim JW, Kim IH, Sonnleitner A, Schrauf M, Curio G, et al. Electrophysiology-based detection of emergency braking intention in real-world driving. J Neural Eng 2014; 11: 056011.

[19] Vinkers CH, Penning R, Hellhammer J, Verster JC, Klaessens JH, Olivier B, et al. The effect of stress on core and peripheral body temperature in humans. Stress 2013; 16: 520-30.

[20] Schulz, P., Kirschbaum, C., Prüßner, J. and Hellhammer, D. Increased free cortisol secretion after awakening in chronically stressed individuals due to work overload. Stress Med.1998;14:91–97.

[21] Hjortskov, N., Garde, A. H., Ørbæk, P. and Hansen, Å. M. Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies. Stress and Health, 2004;20: 91–98. doi:10.1002/smi.1000

[22] Russell E1, Koren G, Rieder M, Van Uum S. Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions. Psychoneuroendocrinology. 2012; 37(5):589-601.

[23] Yalta K, Yalta T, Sivri N, Yetkin E. Copeptin and cardiovascular disease: a review of a novel neurohormone. Int J Cardiol 2013; 167: 1750-9.

[24] Urwyler SA, Schuetz P, Sailer C, Christ-Crain M. Copeptin as a stress marker prior and after a written examination–the CoEXAM study. Stress 2015; 18: 134-7.

[25] Chrousos GP, Gold PW. The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. JAMA. 1992; 267(9):1244-52.

[26] Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983; 24: 385-96.

[27] Monk TH. A visual analogue scale technique to measure global vigor and aff ect. Psychiatry Res 1989; 27: 89-99.

[28] Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.

[29] Whalen KJ, Buchholz SW. The reliability, validity and feasibility of tools used to screen for caregiver burden: a systematic review. JBI Libr Syst Rev. 2009;7:1373-1430.

[30] Aguiló J, Ferrer-Salvans P, García-Rozo A, Armario A, Corbí Á, Cambra FJ, et al. Project ES3: attempting to quantify and measure the level of stress. Rev Neurol. 2015;61:405-15.

[31] Scorrano F, Carrasco J, Pastor-Ciurana J, Belda X, Rami-Bastante A, Bacci ML, et al. Validation of the long-term assessment fhypothalamic-pituitary-adrenal activity in rats using hair orticosterone as a biomarker. FASEB J. 2015;29:859-67.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Stress Management in Primary Caregivers: A Health Challenge

Sira Aguiló[1], Beatriz Lopez-Barbeito[1], Esther García[3], Jordi Aguiló[2,3]
[1]*Hospital Clínic. Barcelona, Spain*
[2]*Department of Microelectronics and Electronic Systems, Universidad Autónoma de Barcelona, Barcelona, Spain*
[3]*Biomedical Research Networking Center (CIBER-BBN), Spain*
acmlis.conference@gmail.com

ABSTRACT

Technology based advances in healthcare are leading to social changes that, in turn, will require innovative responses from health systems. Mainly due to these advances, life expectancy grows so, the aging of world population increases and with it the percentage of chronic patients. Because ofthese two factors: aging people and chronic patients, new needs are posed to the current health system. Together with these patients,a population of non-professional caregivers emerges playing a key role in the care process. However, the impact of the care burden on their health is usually underminedalthough an increasing number of studies report caregiver associated diseases. One of the challenges that arise from health systems: To acknowledgeand to take care this population of caregivers,to avoid the so-called burned caregiver syndrome as well as to reduce the risks of comorbidities associated with the caregiver's stress by setting up appropriate interventions.

Having in mind all these considerations, from the Clinic Hospital in Barcelona we are starting to work ona biofeedback program involvingdifferent populations of caregivers: 30 caregivers of oncological patients and 30 caregivers of geriatric patients and the corresponding control group. The project that is presented in this paper is aiming to determine if these therapies can become effective interventions to reduce the stress that represents the caregiver syndrome. The follow-up of the project willfocus on the possibility to indirectly demonstrate if the decrease of stress actually results in a decrease of the derived diseases and of the total cost that it entails.

Keywords: Primary caregivers, Caregiver's syndrome, Stress Management, Biofeedback

1 Introduction

The increase in life expectancy and in population aging together with medical and healthcare advances are leadingtowards a big change in the world health scenario that will dramatically change the starting point of the public health systems and of course his primary needs to face these new circumstances. The longevity of patients with chronic diseases such as Chronic Obstructive Pulmonary Disease COPD, heart failure or dementia as well as the chronicity of hemato-oncological processes that until a few years ago lacked of treatment have revealed the need to implementsupport units, management and monitoring of *this group of chronic patients.

DOI: 10.14738/tmlai.54.3292
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3292

In parallel, a population of caregivers and primary caregivers of these chronic patients emerges as an indispensable help for the management and accompaniment of this kind of patients while in fact are solving problems and responsibilities of the public health system. Primary caregiver is defined as the individual acting as caregiver of the patient for a minimum of 6 weeks, without receiving financial compensation, with direct emotional involvement and living at home with the patient [1][2]. Primary caregiver is thus subjected to emotional and physical stress derived from the objective burden of the tasks required for patient's care and normally the financing of a significant part of the care expenses and of course the emotional load that all of this represents [3] [4].

In Spain, the economic impact derived from care was1.7-4.9% of Spanish Gross Domestic Product,(GDP) in 2008 [4] while the need for geriatric patient'scaregivers will be about 2.5 million people by 2060 [5]; and this emerging population of caregivers will be subjected to an already recognized physical and psychological stress with its long-term health implications.

From an emotional point of view, usuallycaregivers are not sure about their abilities for the management of the patient,about the best strategy to improve the patient wellbeing or about the better way to achieve theirobjectives [6].Over time, this leads to frustration, anxiety, depression and associated comorbidities such as diabetes, the increase of their cardiovascular risk factor, hypertension or dyslipidemia[7][8] among others which will affect their own health and therefore health and wellbeing of the patient they are in charge of.

From a physiologicalpoint of view, this population of primary caregivers is affected by a deterioration oftheir perceived physical and mental health [9] andoftheir immune system response,they are prone to increase drug consumption[10] and more than half of them will show headache episodes and musculoskeletal problems [11][12]. Moreover, caregivers older than 65 years old who care for their dependent partner have a statistically significant higher mortality risk compared to non-caregivers [13].

The approach to this population of caregivers was initiated years ago, focusing on the collective caring for people in the geriatric age and with a physical and cognitive deterioration.Throughout the last decades some tools have been designed to evaluate the degree of caregiver overload. Among them the Zarit (EZ) Caregiver Overload Scale [14] and The Caregiver Strain Index, which consists of 13 items with yes / no binary responses and where a high level of stress is considered if the caregiver score is greater than 7.

Other initiatives such as the Infosa "Caring for the caregiver" [15] showed that group sessions of empowerment reduced the perception of stress in a population of caregivers of geriatric patients with chronic illnesses.

We are facing a new health challenge: addressing the needs of a growing population of caregivers and proposing preventive measures so that the caregiver's own stress does not lead to long-term illness.The challenge is not straightforward, since so farthere is no consensus about astress measurement method.

Although the WHO has described stress as the new epidemic of s. XXI, several measuring approachesare advancing slowly and still in a very subjective way being the psychometric tests the current gold standard for stress diagnosis and evaluation.

Another approach consideringthe physiological responses of the hypothalamic-pituitary-adrenal system(HPA), and the autonomic nervous system(AN), as well as the different perceptions against the

same or similar stimulus,has high inter-individual variations that make it difficult to establish a quantitative assessment methodorstrategies proposedfor their treatment.

On the other hand, in recent years newprocedures with biofeedback techniques have impacted on the treatment of diseases such as asthma [16], AHT, anxiety or duodenal ulcer [17].Biofeedback techniques consist in giving feedback to the individual with information from hisown physiological variables, such as heart rate, respiratory rate, or muscle contraction force, to know and to learn how these variables can be controledand hence gaininga better control of his own biological responses.In this way, learning and getting control of their own responses we can modulate the AN balance by achieving a greater activity of the parasympathetic system and a reduction of the sympathetic response with the consequent reduction of the physiological cascade generated as response to stress.

In this pilot study, we try to use biofeedback techniques to improve stress perception in a population of primary caregivers of patients with chronic disease and to evaluate their effectiveness according to quality of life indexes or measures, anxiety symptoms, depression and physiological variables such as HRV (heart rate variability) that have demonstrated their involvement in stress [18].

2 Project Objective

The main objective of the project is to determine if biofeed-back techniques are useful as tools to improve quality of life and reduce stress in a population of primary caregivers of chronic patients.

3 Methodology and Experiment's Organization

The experimental pilot study consists in the consecutive inclusion of60 primary caregivers of chronic patients that meet the primary caregiver definition: individual acting as caregiver of the patient for a minimum of 6 weeks, without receiving financial compensation, with direct emotional involvement and living at home with the patient

Two types of primary caregivers will be collected:

On the one hand, caregivers of geriatric patients with chronic disease usually chronic lung disease, heart failure or dementia, who require help for daily life activities due to functional or cognitive impairment.

On the other hand, selection includes caregivers of non-geriatric patients with active onco-hematologic processes, requiring a primary caregiver for agenda compliance, management of side effects of managed treatments and emotional accompaniment according the disease progress.

Included subjects will be cited to a first session, of initial evaluation, where the sociodemographic characteristics, indicated in table 1,and the psychometric tests, indicated in table 2,will be collectedThese tests can be self-administered.An approximate time of 20 minutes is calculated to finish the three tests proposed. However, tests are not time dependent and no fixed response time will be considered for psychometric evaluation.Heart Rate Variability, HRV, trapezoidal Electromyography, EMG, and Respiratory Rate, RR, will be recorded during 10 minutes through a standard electronic equipment NEXUS 10 Mark II, NX10B-EU.

After the initial evaluation session, five biofeedback sessions will follow according to the agreed agenda. These sessions will be devoted to learn and practice about breath control and muscle contraction. Sessions will be spread over five weeks and each lasting about 40 minutes.

The sessions are individual with a therapist trained in biofeedback management.After each sessiona document with detailed information about a series of exercises that should practice at home for 10 minutes a day every day, will be provided. Home practice is to strengthen what was taught in the face-to-face session.

After these five biofeedback sessions, in week 6, the subject will be newly evaluatedfollowing the same procedure than in initial session. That is, the psychometric tests will be repeated, the physical variables will be monitored and a quality of life questionnaire will be passed.

Information of the health status of the patient at week 6 will also be requested and collected regarding their initial state because the patient health statecan influence the stress and quality of life of their caregivers independently to the sessions given.

The results obtained before and after the biofeedback sessions will be compared in order tofindsignificant differences between the degree of stress according to the psychometric tests, the quality of life questionnaireas well as in the physiological signals recorded, HRV, RR, EMG.

4 Material

Continuous monitoring, conditioning and recording of signals will becarried out viaNEXUS 10 Mark II, NX10B-EU de Min Media BV, Netherlands CE mark CE0344 that meetsall the requirements of the MDD (93/42/EEC). MDD Classification IIa and IEC 60601-1-6: 2004 Concerning Electromagnetic compatibility.The system has 12 operating channels in which different sensors can be connected. The main technical characteristics of the measurement device are shown in Table I.

Table 1. Electronic device characteristics

Bipolar Inputs	
Input signaldifference	< 0.1 Vpp
Input commonmoderange	-2.0 V … +2.0 V
Gain factor	19.5
Noise	< 1 µV RMS (1 - 64 Hz freq. range
Input impedance	>= 1012 Ohm (Instrumentation amp at DC)
CMRR (typical)	>= 100 dB
Accuracy	+/- 2 %
Auxiliary Inputs	
Input signalrange	-2.0 V … +2.0 V
Input commonmoderange	-2.0 V … +2.0 V
Gain factor	1
Noise	< 3 µV RMS (1 - 64 Hz freq range)
Input impedance	>= 1010 Ohm (Instrumentation amp at DC)
CMRR (typical)	>= 80 dB
Accuracy	+/- 2 %
AD Conversion	
Resolution	Bipolar 12.2 nV/bit, auxiliary 0.238 µV/bit
Samplefrequency (Bipolar)	8192 Hz, 4096 Hz, 2048 Hz, 1024 Hz, 512 Hz, 256 Hz, 128 Hz
Samplefrequency (auxiliary)	256 Hz, 128 Hz
Channelbandwidth	DC up to 0.4 * sample frequency

The device is organized in 12 operating channels in which sensors could be connected. Resolutions and signal ranges are different depending on the channel number. There are 4 dedicated channels. Table II summarizes Chanel function, range and resolution.

ChanelNr.	Function	Resolution/bit	Signalrange
1, 2, 3, 4	Bipolar signals	12.2 nV	±100 mV
5, 6, 7, 8	Auxiliarysignals	0.24 µV	±2.0 V
9	Oxygensaturation	1 %	0-100 %
10	Plethysmographic	1 (bit)	0-255
11	Pulseoximeterheartrate	1 BPM	0-255 BPM
12	Pulseoximeter status	1 (bit)	0-255

5 Discussion

Caregivers of chronic patients is an emerging population, with a role not always well defined, submitted to physical and emotional stress.Better management focusing on the stress reduction is important in mitigating long-term health consequences such as anxiety, depression, musculoskeletal problems or cardiovascular events, and to ensure good care of those who are in charge.

Although different subpopulations of caregivers have not yet been defined, it seems obvious that that there will be different groups of caregivers with different needs. As an example, it seems to be clear that caregivers of geriatric patients with cognitive impairment needing help for basic activities such as grooming or feeding are a population other than caregivers of autonomous patients diagnosed with an active hemato-oncological process.

The current health challenge in this field is in the context of an increasingly aging population, with chronic diseases and active onco-hematological processes that last over time.

This implies the need for a conceptual change, but also a practical one allowingto takecare of the needs of this group of patients, as well as of those of their caregivers, who are indispensable pieces of health processes.

It is well established that being a primary caregiver is stressful and that it has long-term consequences. However, interventions are difficult because of the interindividual differences in the perception and response to stress and the difficulty of not having a validated tool to quantify and / or measure the stress level.

Given the already stated sociodemographic evolution, to propose and to test strategies for improvement and control of the caregiver's stress and of its consequences should be a major health objective.

Biofeedback techniques seem to be a good tool for stress management and therefore, to use it as therapeutic strategy to improve quality of life in this population group. Once it is clear from publications [19] and from our own experience that the stress level could be quantitatively evaluated, what we are doing at the Clinic Hospital in Barcelona presented in this paper is to prove that biofeedback techniques could be an effective treatment to reduce the stress level. Additionally, to prove the efficacy of the applied techniques we will measure the quality of life,the QOL Index [20], using scales of perceived stress and using also the physiological variables that are constituting a stress biomarker.

Although there are different studies approaching problems of the caregiver's population, most of them are merely descriptive and don't propose improvementstrategies or evaluate possible effects of the implementation of this kind of interventions in this group.

The characterization of the behavior of both groups of caregivers will allow to observe the different response to the proposed interventions. Which will allow the modeling of the system and offer customized interventions

ACKNOWLEDGEMENTS

This research line is supported by the Hospital Clinic de Barcelona, http://www.hospitalclinic.org/, the CIBER-BBN (http://www.ciber-bbn.es)and the Univ. Autónoma de Barcelona throughthe MINECO (FIS-PI12/00514 and TIN2014-53567-R)

REFERENCES

[1] [1] Mateo, I, Millán, A, García, M, Gutiérrez, P, Gonzalo, E. y López, L. A." Cuidadores familiares de personas con enfermedad neurodegenerativa: perfil, aportaciones e impacto familiar", Atención Primaria. 2000; 26; 25-34.

[2] Rodríguez, P., Sancho, T., Alvaro, M. y Justel, M." Las personas mayores en España. Perfiles. Reciprocidad familiar. Madrid: Instituto Nacional de Servicios Sociales

[3] Pinquart M, Sörensen S. "Associations of stressors and uplifts of caregiving with caregiver burden and depressive mood: a meta-analysis". J Gerontol B Psychol Sci Soc Sci. 2003;58:P112-28.

[4] Oliva-Moreno J, Trapero-Bertran M, Peña-Longobardo LM, Del Pozo-Rubio R. "The Valuation of Informal Care in Cost-of-Illness Studies: A Systematic Review". Pharmacoeconomics. 2017; 35: 331-345..

[5] Pickard L, King D. Informal care supply and demand in Europe. In: Geerts. J, Willeme P, Mot E, editors. Long-term care use and supply in Europe: projections for Germany, the Netherlands, Spain and Poland.Brussels: European Network of Economic Policy Research Institutes (ENEPRI). 2012.Research Report 116:114–23.

[6] Lyons JG, Cauley JA, Fredman L. The Effect of transitions in caregiving status and intensivity on perceived stress among 992 female caregivers and noncaregivers. J Gerontol A Biol Sci Med Sci. 2015 Aug;70(8):1018-23.

[7] Susan K.Roepke, Matthew Allison, Roland Von Ka Nel, Brent T. Mausbach, et al. "Relationship between chronic stress and carotid intima-media thickness (IMT) in elderly Alzheimer's disease caregivers". Stress, March 2012; 15(2): 121–129.

[8] Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries. (the INTERHEART study): case-control study. The Lancet; Sep 11-Sep 17, 2004; 364, 9438; Health & Medical Collection.

[9] Haug, M. R., Ford, A. B., Stange, K. C., Noelker, L. S., & Gaines, A. D. (1999). Effect of giving care on caregivers' health. Research on Aging, 21(4), 515-538.

[10] Janice K. Kiecolt-Glaser, Lymama Mc Guire, Theodore F Robles, Ronald Glaser. Psychoneuroimmunology: Psychological Influences on Immune Function and Health. Journal of Consulting and Clinical Psychology 2002; vol 70; 537-547.

[11] Ashi, S., Girona, G. y Ortuño, M. A. (1995). Sindrome del cuidador. Rehabilitación, 29, 465-468.

[12] M.J. Gómez-Ramos, F.M. Gonzáñez-Valverde. Caregivers of dementia patients: application of the Caregiver Burden Scale. Rev Esp Geriatr Gerontol 2004;39:154-9.

[13] Schulz R, Beach SR. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. JAMA. 1999;282:2215-9.

[14] Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. Gerontologist. 1980;20:649-55.

[15] http://blog.hospitalclinic.org/es/2014/06/presentacion-del-programa-infosa-de-apoyo-a-cuidadores-de-personas-dependientes/.

[16] Cheryl L. Kern-Buell, Angele Mc Grady, Philip B. Conran, Lois A. Nelson. Asthma severity, Psichophysiological indicators of arousal, and inmmune function in asthma patients undergoing biofeedback-assisted relaxation. Applied Psychophysiology and biofeedback. June 2000, vol 25. Issue 2, 79-91.

[17] Juan Manuel Vázquez Goñi et al, "Reducing the stress of food processing workers through respiratory biofeedback". Revista Cubana de salud y trabajo 2011; 12 (1): 39-44.

[18] Hernando A, Lazaro J, Gil E, Arza A, Garzón-Rey JM et al. Inclusion of respiratory frequency information in Heart Rate Variability for a better analysis of Stress.IEEE JBHI2016; 20 (4): 1016-1025

[19] Aguilo J, Ferrer-Salvans P, Garcia-Rozo A, et al. Project ES3: attempting to quantify and measure the level of stress. Rev Neurol. 2015;61(9):405-415.

[20] Dunn SA Bonner PN, Lewis SL, Meize-Grochowski R. Quality of life for spouses of CAPD patients. ANNA Journal. 1994;21:237–247.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Remote Assessment of Disease and Relapse– Central Nervous System-RADAR-CNS

J.M. Garzón-Rey and J. Aguilo
Department of Microelectronics and Electronic Systems, Universidad Autónoma de Barcelona, Barcelona, Spain
Biomonitoring Group, Biomedical Research Networking Center (CIBER-BBN), Spain
acmlis.conference@gmail.com

ABSTRACT

Remote Assessment of Disease and Relapse in Central Nervous System Disorders (RADAR-CNS) is a major international research project. It aims to develop new ways of measuring major depressive disorder, epilepsy and multiple sclerosis (MS) using wearable devices and smartphone technology. RADAR-CNS aims to improve people's quality of life and change how depression, epilepsy and MS are managed and treated. Data from mobile devices can give a full picture of a person's condition at a level of detail which was previously impossible. This offers the potential to detect changes in behavior, sleep, or mood before the individual themselves is aware of it. This could help them to predict or even avoid a relapse. To achieve this, we are creating a pipeline for developing, testing and implementing remote measurement technologies for depression, multiple sclerosis (MS) and epilepsy. Depression, multiple sclerosis and epilepsy are all disorders of the central nervous system. While the symptoms and disability experienced by individuals with each condition are different, they all have a significant effect on people's wellbeing. For doctors and people with these long-term or chronic conditions, understanding how their disease changes over time can help with its management. But in chronic conditions most of the symptoms and episodes happen outside of the health care environment. Measuring individuals' symptoms, mood and daily function continuously, could help people gain better insight into their condition. RADAR-CNS receives funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA.

Keywords: Central Nervous System Disorders, Depression, Multiple sclerosis and epilepsy

1 Introduction

Remote Assessment of Disease and Relapse in Central Nervous System Disorders (RADAR-CNS) is a major international research project. It aims to develop new ways of measuring major depressive disorder, epilepsy and multiple sclerosis (MS) using wearable devices and smartphone technology.

RADAR-CNS aims to improve people's quality of life and change how depression, epilepsy and MS are managed and treated. Data from mobile devices can give a full picture of a person's condition at a level of detail which was previously impossible. This offers the potential to detect changes in behavior, sleep,

DOI: 10.14738/tmlai.54.3293
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3293

or mood before the individual themselves is aware of it. This could help them to predict – or even avoid – a relapse.

RADAR-CNS brings together clinicians, researchers, engineers, computer scientists and bioinformaticians from all over the world. It is jointly led by King's College London and Janssen Pharmaceutical NV. The project is funded by the Innovative Medicines Initiative a Public Private Partnership set up between the European Federation of Pharmaceutical Industries and Associations (EFPIA) and the European Union. It includes 23 organizations from across Europe and the US.

2 Strategy

RADAR-CNS aims to transform patient care through remote assessment using wearable technologies such as smartphones or fitness trackers. The project will develop the technology to identify which clinical or physiological biosignatures can be measured remotely to predict relapse or deterioration.

To achieve this, we are creating a pipeline for developing, testing and implementing remote measurement technologies for depression, multiple sclerosis and epilepsy. Patients will be involved at each stage of development. It will include a generic data management and modelling infrastructure applicable to other mental and physical disorders. It will be designed with the flexibility to be sustainable alongside future technological developments.

We will anticipate and solve any potential problems with using these technologies by consulting patients, clinicians, regulators and healthcare providers throughout the project.

We bring together an international consortium of academic and EFPIA members who are leaders in the fields of depression, multiple sclerosis and epilepsy. We also deliver clinical expertise and access to patient cohorts in each disease area. This is combined with leading technical and methodological expertise in the disciplines required to develop and implement remote measurement technologies.

3 Conditions

Depression, multiple sclerosis and epilepsy are all disorders of the central nervous system. While the symptoms and disability experienced by individuals with each condition are different, they all have a significant effect on people's wellbeing. For doctors and people with these long-term or chronic conditions, understanding how their disease changes over time can help with its management. But in chronic conditions most of the symptoms and episodes happen outside of the health care environment. Measuring individuals' symptoms, mood and daily function continuously, could help people gain better insight into their condition.

3.1 Major Depressive Disorder

Major depressive disorder, sometimes called "clinical depression" or "depression", can be triggered by a life event, or result from stress, or happen without a specific cause. It is the most severe form of depression where people exhibit a sense of hopelessness and despair along with low mood and negative thoughts. This can affect the way people eat, sleep, feel about themselves, and think about things. Without treatment, the symptoms can last for weeks, months, or even years.

People with depression can lose interest in activities they once enjoyed, withdraw socially, and have trouble concentrating or making decisions, and may experience changes in eating and sleeping patterns.

Physical symptoms of depression can include fatigue, weight change, a feeling of being slowed down, headaches, digestive disorders, or chronic pain.

At its worst, people with depression can feel like they want to harm themselves, and experience or act on thoughts of suicide. There are several different types of depression, and RADAR-CNS focuses on Major Depressive Disorder, which affects approximately 7% of adults in Europe. Depression is thought to affect over 30 million people in the European Union [1].

3.2 Multiple Sclerosis

Multiple sclerosis is a condition in which the body's immune system has an abnormal reaction to the central nervous system (CNS, made up of the brain, spinal cord and optic nerves). The immune system attacks elements of the CNS, disrupting the transmission of nerve signals and producing the symptoms that characterize MS. The exact cause is unknown, but is thought to be triggered by a combination of environmental and genetic factors.

MS can be mild, moderate or severe, and can show different patterns of symptoms over time. Most people (85%) with the disease have the relapsing remitting form, which involves clearly defined attacks of neurological symptoms followed by periods of partial or complete recovery. Some people experience progressive MS where symptoms and disability worsen over time. For many people with MS, as time goes on their symptoms do not resolve completely, resulting in gradually worsening illness and disability.

Treatments for MS involve disease modifying drugs which slow the progression of the disease and reduce the frequency of the attacks. There is no cure for MS and the challenge lies in managing the disease as best possible to continue to live a normal life. MS affects around 700000 people in Europe.[2], [3]

3.3 Epilepsy

Epilepsy is one of the most common serious neurological conditions in the world, affecting around 6 million people in Europe. Epilepsy affects the brain, causing seizures. A seizure is caused by a sudden burst of intense electrical activity in the brain, temporarily disrupting its normal function.There are many different seizure types. In some seizures the person remains alert, but experiences altered sensations and perceptions. In others they lose consciousness, and their muscles may stiffen and jerk.

Epilepsy is usually diagnosed after someone has more than one seizure, and there is considered to be a high probability that they will have further seizures. Some types of epilepsy last for a limited time and the person eventually stops having seizures, but for most people, epilepsy is a life-long condition.

Epilepsy has no cure, but taking anti-epileptic drugs (AEDs) helps people to prevent or reduce the frequency of seizures. AEDs do not work for everyone, so other treatments may include brain surgery, vagus nerve stimulation or a special diet.[4], [5]

4 Governance & Funding

The Figure. 1 shows the overall organizational structure of RADAR-CNS, and the bodies responsible for specific aspects of project management.

To ensure that the views of people with conditions are incorporated into RADAR-CNS we have set up a Patient Advisory Board that includes people with depression, epilepsy and multiple sclerosis along with representatives from relevant support organizations. Their role is to provide feedback on the layout and

content of research materials and provide expert opinions on important decisions to be made in the design of studies.

RADAR-CNSis a large and complex project, involving collaborators from diverse disciplines and 23 partner organizations from across Europe and the US. To ensure the project is conducted efficiently and thoroughly, the project's research program is arranged into 11 work packages. Work packages are arranged into three clusters – Clinical Disorders, Translational Pathways, and Technical Platforms – with the Central Management cluster providing support for the project as a whole. Membership between work packages is strongly overlapping, and many centers and investigators contribute to multiple work packages.

4.1 Central Management

This cluster provides project management, data management and communications support for the RADAR-CNS project as a whole.

Project Management (Work Package 1)

Provides overall management of the project, ensures that its contractual duties are carried out, and establishes and maintains effective communication within the project.

Dissemination, Exploitation and Communication (Work Package 11)Deals with the dissemination, communication and exploitation of the results generated in RADAR-CNS. This includes raising public and scientific awareness of the project and managing processes for the capture and protection of intellectual property.

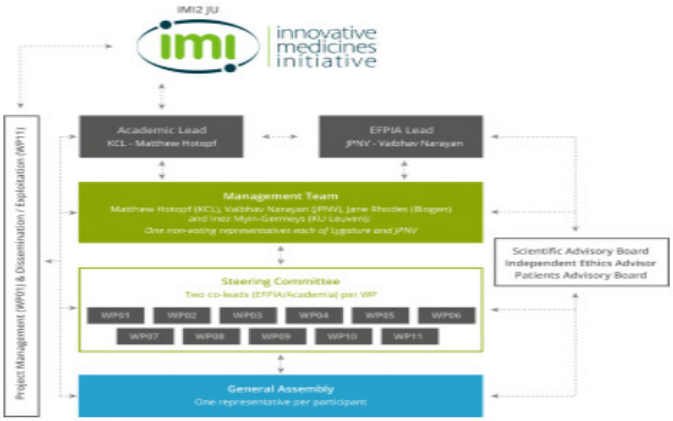


Figure 1overall organizational structure of RADAR-CNS, and the bodies responsible for specific aspects of project management.[7]

4.2 Clinical Disorders

This cluster brings together clinical expertise across RADAR-CNS relating to our three focus conditions: depression, epilepsy and multiple sclerosis. This cluster focuses on evaluating the feasibility, adherence and personal satisfaction with remote measurement technologies in people with different conditions. It also addresses the clinical harmonization required to assess sleep, physical activity, speech, mood, and cognition in everyday life across clinical disorders.

4.2.1 Major Depressive Disorder (Work Package 6)

At the heart of this work package is a multicenter observational cohort study following patients with recurrent depressive disorder. Participants will use remote measurement technologies whilst being followed for clinical outcomes (particularly, though not exclusively, recurrence and relapse) over the course of 12 months. This study will be complemented by learning from the ORBIT [6]and, although not identical in design, the two studies will be sufficiently close to allow us to test whether findings from one generalize to the other. The project lead of the Orbit study is a participant of this work package and will provide the full support and experience required.

4.2.2 Epilepsy (Work Package 4)

We are initiating a set of studies into remote measurement technologies across a range of parameters. We aim to determine the feasibility, acceptability of, and adherence to, remote measurement technologies in people with epilepsy to provide real-time objective, multidimensional indications of seizure occurrence and clinical state in epilepsy. We are also exploring whether the trajectory of non-seizure-related aspects (e.g. stress, mood) associate with, and predict/anticipate, changes in the trajectory of frequency of seizure occurrence.

4.2.3 Multiple Sclerosis (Work Package 5)

Exploring whether remote measurement technologies can be used to characterize two important features of clinical presentation, depression and gait disturbance in people with multiple sclerosis. We are determining whether remote measurement technologies can detect mood changes in MS patients in the early stages of their condition, and exploring whether they can reliably assess the clinical disability of MS patients. We are also investigating these technologies' sensitivity in detecting state changes and disability progression over time, evaluating the interaction among fatigue, depression and sleep disorder and sensory-motor dysfunction, including gait.

4.2.4 Clinical Harmonisation (Work Package 3)

Providing an overarching assessment and analysis scheme to assess sleep, physical activity, speech, mood, and cognition in everyday life across clinical disorders. We are focusing on variability in sleep quality, levels of activity, social interactions, mood, cognitive performance and stress as possible predictors of clinical course. We are also examining how remote measurement technologies can monitor and improve quality of life and psychological well-being for people with depression, epilepsy, or multiple sclerosis.

4.3 Technical Platforms

This cluster provides the technical and computing knowledge required to bring the RADAR-CNS project to clinics. As well as producing a flexible, reusable platform that can be used across numerous conditions and scenarios, the cluster also provides the data analysis expertise required to understand the data collected and how they relate to clinical outcomes.

4.3.1 Devices and Platform (Work Package 7)

Building an end-to-end system to support passive and active remote measurement and feedback using mobile and web technologies. We expect emerging monitoring technologies to continue to evolve rapidly at unpredictable rates. As such, the generic architecture of the RADAR platform is intended to assist incorporation of new technologies in the short term for RADAR-CNS, but will also form a reusable, flexible

architecture that can power future RADAR projects (e.g. RADAR-DIABETES). A technology steering committee will ensure the close working between the clinical, patient engagement and technology Work Packages.

4.3.2 Data Analysis & Biosignatures (Work Package 8)

Enabling data collected as part of the project to be analyzed. We aim to understand the association between these data and the remission, relapse, and recurrence of the conditions being studied, and to classify and predict the disease status. Requires the creation and refinement of algorithms for data extraction and filtering; analysis of the collected data to create models of diseases and its fluctuations; and finally the generation of data to enable patient and clinician engagement tools.

4.4 Translational Pathways

This cluster ensures that findings from the RADAR-CNS Clinical Disorders and Technical Platforms clusters are translated efficiently and effectively into real-world clinical applications. We aim to have workable clinical devices entering health services by 2020. To achieve this, we need to understand and learn from the needs of patients, clinical services and regulators.

4.4.1 Patient Involvement (Work Package 2)

Works with patient stakeholders to understand issues such as privacy, usability, and acceptability of remote measurement technologies. Also works to identify clinical endpoints which are most relevant to patients, and identify facilitators and challenges for engagement and adherence that can be further tested in this project.

4.4.2 Clinical Pathways (Work Package 9)

Identifies the requirements of clinical stakeholders (healthcare professionals, managers, commissioners and payers) for integrating remote measurement technologies into care pathways for depression, epilepsy, and multiple sclerosis in different European healthcare systems. Provides a structure so that these requirements form a coherent contribution to the project system specification to ensure its maximum applicability, acceptability and adoption in healthcare systems.

4.4.3 Regulation (Work Package 10)

Works with EU medical device and product regulators to understand their requirements in this field. Provides a structure to ensure that the regulatory learning during the project is collected and integrated to ensure its maximum applicability, acceptability and adoption. This work package will provide a framework for the classification of remote measurement technologies as medical devices or otherwise, and will also identify what kind of information is needed for regulatory approval.

ACKNOWLEDGMENT

RADAR-CNS receives funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. See http://www.imi.europa.eu/ for more details

REFERENCES

[1] H. U. Wittchen, F. Jacobi, J. Rehm, A. Gustavsson, M. Svensson, B. Jönsson, J. Olesen, C. Allgulander, J. Alonso, C. Faravelli, L. Fratiglioni, P. Jennum, R. Lieb, A. Maercker, J. van Os, M. Preisig, L. Salvador-Carulla, R. Simon, and H.-C. Steinhausen, "The size and burden of mental disorders and other disorders of the brain in Europe 2010," Eur. Neuropsychopharmacol., vol. 21, no. 9, pp. 655–679, 2011.

[2] "MS Info, Research, News and Support | Multiple Sclerosis Society UK." [Online]. Available: https://www.mssociety.org.uk/. [Accessed: 03-Apr-2017].

[3] "Home : National Multiple Sclerosis Society." [Online]. Available: http://www.nationalmssociety.org/. [Accessed: 04-Mar-2017].

[4] "Epilepsy Action | Epilepsy information and advice." [Online]. Available: https://www.epilepsy.org.uk/. [Accessed: 04-Mar-2017].

[5] "Epilepsy Information, Support & Research | Epilepsy Society." [Online]. Available: https://www.epilepsysociety.org.uk/. [Accessed: 04-Mar-2017].

[6] J. P. Piccini, D. N. Simon, B. A. Steinberg, L. Thomas, L. A. Allen, G. C. Fonarow, B. Gersh, E. Hylek, P. R. Kowey, J. A. Reiffel, G. V. Naccarelli, P. S. Chan, J. A. Spertus, and E. D. Peterson, "Differences in Clinical and Functional Outcomes of Atrial Fibrillation in Women and Men," JAMA Cardiol., vol. 1, no. 3, p. 282, Jun. 2016.

[7] "RADAR-CNS | RADAR-CNS." [Online]. Available: http://www.radar-cns.org/. [Accessed: 04-Mar-2017].

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Conception and Evaluation of Internet of Things Connecting to RFID Framework

[1]M. El Harrak Hajri, [1]A Asselman, [2]M. Yassin Chkouri
[1]Physic Departement, Science Faculty of UAE, Tetouan, Morroco
[2]Informatic Departement, ENSA Tetouan of UAE, Tetouan, Morroco,
harrakmed@gmail.com, adelasselman@gmail.com, yassin.chkouri@gmail.com

ABSTRACT

The notion of the Internet of the Objects designates a network of objects connected to each other and communicating automatically. In this paper we develop the notions of semantics are becoming increasingly important because, more than ever, they appear as a solution to problems of interoperability and interpretation of data and services by machines. The diversity of possible applications at the intersection of the Internet of Things and the Semantic Web has prompted many research teams to work at the interface between these two disciplines. We wish in this paper to make an inventory of their proposals. We also seek to contribute to the evolution of this field of research by proposing an ontology to describe networks of connected objects. Keywords: Ontologies, Internet of Objects, Interoperability, Data Enrichment.

Keywords—RFID ;Internet of things ;network topology ; microcontrolles

1 Introduction

In recent years, the Internet of objects (IoT in English 1) has evolved at an exceptional speed, connecting a large number of heterogeneous objects (sensors, actuators, smartphones, applications, etc.). The IoT contributes significantly to the production of data under Big Data, with an estimate of 4.4 billion GB produced in 2013, with some estimates projecting the existence of more than 26 billion IoT devices connected to the " Horizon 2020 2. One of the main concepts of the internet of objects is machine-to-machine (M2M). M2M is the association of information and communication technologies, with communicating machines in order to provide the latter with the means to interact without human intervention. The fields of application are thus very wide: energy management, home automation, medical diagnostic aid, comfort of life, etc. M2M has suffered from a vertical fragmentation of the approaches adopted to cover the needs of the different fields of application. To solve this problem, in 2012 the European Standardization Institute (ETSI) developed a standard specification for a horizontal M2M service platform, the first of which was provided by LAAS. This makes it possible to collect data and manage various objects in a transparent way. The challenge now lies in the use of these data. The best approach seems to be to associate semantics to these data through ontologies in order to facilitate their reuse and to allow the implementation of the reasoning mechanisms. Recent works aim to propose ontologies allowing to represent the data collected by connected objects as well as the interactions between these objects Barnaghi et al. (2012), L. Atzori & Morabito (2010). These approaches are, for the

DOI: 10.14738/tmlai.54.3320
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3320

most part, specific to the types of objects for which they are designed and to the data they collect. A real challenge is to define a generic approach for the semantization of data with a dual objective: to promote interoperability between objects, and to make data usable automatically. They could then be integrated into platforms that implement automatic diagnostic and supervisory mechanisms, which is crucial for M2M. The purpose of this paper is twofold. We first present a panorama of the needs that ontologies can meet in the field of IoT and present the ontologies currently available in the field. On the basis of this analysis, we propose an ontology to capitalize on the one hand the work done in the field and on the other hand to take into account the mature advances in the field of Knowledge Engineering. The paper is organized in the following way: Section 2 presents a state of the art in the use of ontologies in IoT, Section 3 describes the IoT-O ontology that we propose and Section 4 illustrates its interest From a real case[1].

2 Concepts and Definitions

2.1 The Internet of Things

The internet of objects is a technological revolution in the field of computing and telecommunications [2, 3]. IoT refers to a variety of detection equipment and information systems such as sensor arrays, reading devices (RFID, barcode), short-range location and communication systems based on Machine-to-machine (M2M) communication, through the Internet to form a larger and smarter network [4]. This revolution is based on a constant evolution of the Internet, technologies and software, communication protocols, embedded sensors that are constantly being improved, physical objects that are increasingly intelligent and capable of providing information and To perceive their environment in real time [8]. The IoT can be seen from two angles, either centered on the Internet or centered on the object. When it is centered on the Internet, services are the main component of its architecture and the objects contribute by feeding it with data. When centered on the object, the center of the architecture becomes the object and one speaks of the cloud of the objects. The cloud of objects thus appears as a platform of objects allowing intelligent use of infrastructures, applications and information at reduced costs [5].

IoT is based on a wide range of technologies, protocols, networks and concepts: network infrastructures, new software, hardware and services platforms. In particular, IoT is associated with identification and traceability via the integration of RFID (Radio Frequency Identification Systems) chips; The semantic web, nanotechnologies, mobility, ubiquity, crowdfunding [6]. This constant and evolving change in technologies and the advent of new platforms, new services and new architectures bring new prospects, new markets with wide, varied economic, social, political, ethical, security and regulatory challenges.

We can cite as challenges the integration and sharing of data on Cloud platforms, securing personal data of users (freedom and confidentiality), good governance (transparent and democratic), harmonization of standards, networks And the vagaries of economic competition [7]. The management of the healthcare chain is in this perspective and occupies a preponderant place among the fields of application of these new technologies and concepts related to the IoT. We can cite as examples the problems dealt with in the field of healthcare, as discussed in [8], such as management and decision support, stock optimization, Quality of service, radio frequency identification, real-time tracking of products and processes.

2.2 The health chain

Sanitation can be seen as the integration and collaboration of a set of activities whose purpose is to plan, implement and control a flow of materials, semi-finished objects and finished projects, Origin at point of

use. These activities include planning and forecasting, control and storage, handling, order processing and distribution, marketing, purchasing, packaging, after-sales service [9]. All these activities result in product-related information flows and financial flows, the optimization of which is a key factor for the competitiveness of companies. Controlling flows (physical flow and information flow) makes it possible to make strategic and tactical decisions for the good governance of the company according to the specific constraints that may be financial (inventory immobilization costs, outstandings), physical (Warehouse location) or environmental (reduction of pollution, waste and energy consumption) [10].

Many tools and software have been created to facilitate the control of these flows. These tools include Enterprise Resource Planning (ERP), Advanced Planning System (APS), WareHouse Management System (WMS), Transport Management System (TMS), Customer Relationship Management (CRM) [11].

Most of these tools are struggling to adapt to the new challenges of the current health chain, namely demand uncertainty, coordination and global governance of the health chain [12], An intelligent and dynamic control [13], the flexibility of the chain to better meet the demands of the customer [14]. Similarly, these tools must deal with risk management and local and global decision-making for a decentralized health chain [15]; The interoperability of health networks with the constraints of heterogeneity of norms and standards. The management and exploitation of the flow of information generated by networks and healthcare platforms as an added value for the competitiveness of companies and as a vector for the creation of new structures and markets ("virtual enterprises", "enterprise 2.0") Represent challenges to be addressed by the aforementioned tools [16]. The Internet of objects as presented above contributes to the resolution of some of these great challenges. We cite as an example the identification by radio frequencies of products and sanitary objects (RFID, NFC (Near Fields Communication)) whose uses are multiple: traceability of products, management of the production chain, management of subscribers in Transport and leisure,

3 The Internet Technologies of Things

The Internet of objects aims to connect objects to each other via Internet protocols. The THING here represents everything around us (machines, mobile phones, computers, sensors) [17]. To achieve this goal, it is imperative to be able to identify the objects, assigning them a virtual interface so that they can communicate with their environment. It is important to note that more than six billion objects will be connected by 2015 [18]. The technological domains covered by the IoT are broad and varied, as illustrated by the deployment diagram in Figure 3.

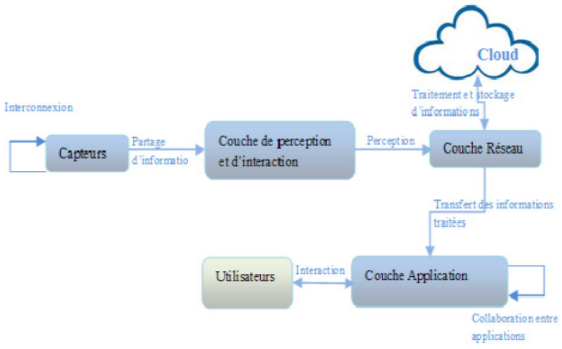


Figure 1: Deployment diagram in the IoT [19].

In the following, we present a non-exhaustive list of the different Io concepts and technologies and their applications in the field of health.

3.1 Architectures, Protocols and Algorithms

New models of architecture make it possible to integrate sensors and the Internet. This communication between a sensor and the Cloud is done by a virtual layer that implements the operation of the actual sensors. Such a layer gives rise to new architectures called sensor arrays (Sensor Cloud) or Cloud of virtual sensors [20]. With such an architecture it is possible to create services based on virtual sensors, that is to say geographically distributed sensor environments which can be used on demand by several users.

In the field of sanitation, and in particular in the management of wind farms, such architectures are used for the transmission of data for preventive maintenance purposes. In this area, many other applications are underway to link mathematical optimization algorithms to web services to better harness the data collected by the sensors in order to provide better tools and maintenance services [19]. With the IoT, billions of objects will be connected to the Internet, each object possibly possessing its own sensors. This ecosystem poses problems in identifying and locating objects and services (sensor, actuators, pallets, container, service, etc.) uniquely in the Internet, and makes it even more difficult to manage the enormous volume Of information generated by these sensor networks.

To address this issue, several protocols and algorithms have emerged, including CASSARAM, a middleware offering semantic-based discovery models; The Objects Naming Service (ONS) to address

In order to facilitate the search and the identification of the objects, to improve the communication between the objects and the platforms of the Cloud. The ONS is a naming service that distributes information about the source of a product, from the manufacturer to the consumer. It is based on the same principle as the DNS (Domain Name System), to access it simply by knowing the product identifier EPC (Electronic Product Code), or the GTIN (Global Trade Item Number) [12]. Several architectures (2-Tiers, 3-Tiers) linking the production chain to the IoT have been proposed in order to set up manufacturing platforms based on Cloud Computing [15]. Tools for driving the production line from RFID tags, as is the range of IFM products used in the production chain. An application of these new architectures based on the IO in healthcare is the production chain where it is a question of sharing the resources (machines, robots) and the production capacities optimally, and make the resource allocation to the [17]. Another feature of this application is the ability to use these protocols to inform the customer about the origin and content of the product. By integrating sensor data with Cloud platforms, IoT architectures provide wider use possibilities that extend beyond the sanitary sphere.

3.2 Interfacing technologies (RFID, NFC, Zigbee) in the IoT

Several technologies are used to communicate an object with the Internet, including RFID, NFC, Zigbee communication protocol [19]. Other solutions are being developed such as acoustic identification systems, microwaves, optical systems, the use of DNA, software marking or the integration of chips in the design of objects [18]. The RFID consists of a read / tag pair. The reader sends a radio wave, the label in turn sends an identification frame. Once the chip is fed, the tag and tag communicate according to the Tag Talk First (ITF) or ITF (Interogator Talk First) protocol. In TTF mode, the tag first transmits the information contained in the chip to the interrogator. In ITF mode, the interrogator sends a request to the tag, and the tag responds later. There are three types of labels, passive, active and semi-active labels. The former do not have their own source of energy: a small amount of energy is provided by the magnetic field

induced by the reader at the time of identification. The active tags are battery-powered, they are capable of sending themselves identification information without soliciting a reader. The semi-active uses a hybrid mechanism: self-powered, they activate only at the request of the reader, allowing a lower energy consumption than the active tags. The reading distance of the RFID chips varies from a few cm to a few meters (10 m), and can go beyond (200 m) with long-range communication technologies [19]. The operating principle of the RFID technology is illustrated in Figure 5.

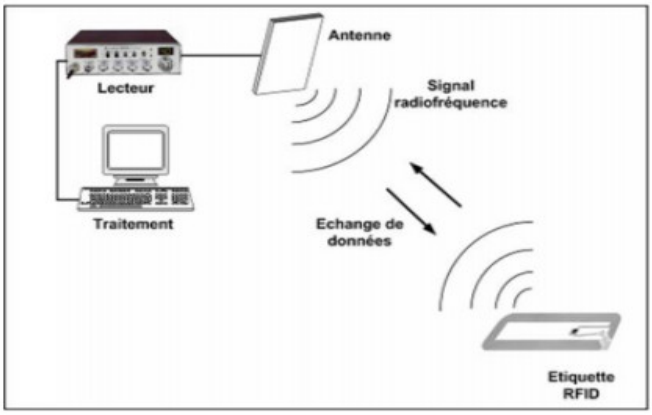


Figure 2: Principle of operation of RFID

NFC technology is the result of several evolutions of microcontrollers, smart cards, and short-range communications [11]. NFC is based on the same principle as RFID, ie identification by radio frequency. It allows the exchange of information at short distance between two objects (a reader and a card) without contact, and operates in two modes: passive mode and active mode. In passive mode, the user's terminal emulates a smart card and acquires energy from the reader's radiation (mobile phone, for example). In active mode, the terminal behaves like an electronic label reader (barcode, 2D labels) and has its own power source [20] (an on-board battery for example). NFC allows the user to exchange information with his environment, in particular in the field of transport, leisure, shopping, or reading information on billboards. The use of NFC facilitates the management of sales data in the health chain, management and validation of bus tickets in urban transport [11]. When a user scans an NFC tag, they can also access information about the product (origin, manufacturer, content / ingredients, manufacturing process) [11].

Zigbee is a low-cost wireless communication protocol that allows short-distance exchanges between nodes on a Wireless Personal Area Networks (WPAN) network. This protocol is based on the IEEE 802.15.4 standard, which specifies the communication protocols between the physical layers and the data link of the OSI model, by defining three types of equipment: Full Function Devices (FFDs) Complete, reduced Function Devices (RFDs), and network coordinators. The FFDs coordinate the entire network, they are Personal Area Network (PAN) coordinators, router or device connected to a sensor. RFDs are designed for simple applications such as lighting a lamp. RFDs can communicate only with an FFD [49]. Among the advantages of this communication protocol, we can mention the low energy consumption, the optimum use of the bandwidth, and its low implementation cost. These advantages make it possible to adopt the Zigbee protocol in embedded environments and industrial networks, or the development of new products based on this protocol [12].

Figure 3: XBee Communication Module

XBee can be used to transmit data between sanitary objects (sensors on containers, RFID tags on products) and cloud environments (data processing, Big Data, Web services).

3.3 Communication networks (GSM, UMTS, LTE / LTE-A, WSN, VANET) in the IoT

With the advent of the IoT, another challenge is the availability of reliable communication networks, both in terms of infrastructure and communication protocols. This challenge is linked to, among other things, the mobility of objects, the heterogeneity of data and platforms, access to information from any place, anytime and anywhere Device (PDA, Smartphone, tablet), which makes it more difficult to standardize protocols and algorithms. Moreover, the volume of data to be transmitted by the sensors places great stress on the availability of the bandwidth, hence the need to implement networks adapted to these new constraints. In the literature, mention is made of two approaches, the use of short-range wireless networks (Zigbee, WiFi) to connect objects to the Internet via a gateway; And conventional broadband cellular networks (4G, 3G). The use of Ultra Narrow Band (UNB) networks is booming because most objects do not require the bandwidth available in broadband networks, but rather Networks with very low costs and very low energy consumption. Among the players in these ultra-low-band networks, we can cite Sigfox, Neul and On-Ramp. Sigfox networks are characterized by their hierarchical structure; Servers that verify all data and route messages to information systems, UNB modems that communicate with base stations or cells to cover wide areas, and base stations that route messages to servers . Other solutions exist to address communication problems in IO, namely the integration of ubiquitous wireless networks to manage object connectivity [18], virtualization of network resources (physical and virtual) to facilitate Sharing and availability of resources, Xbee networks [14], the use of wireless sensors, for home automation applications, environmental monitoring, and portal security. These technologies are used in urban transport sanitation for vehicle collision avoidance [15], the location of people in risk areas by the use of wireless sensors, the prediction of transport time for goods, Or in the medical facility to collect information during an organ transport [16].

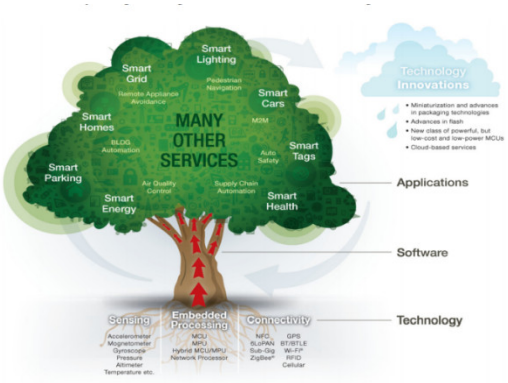


Figure4: the IOT - Different Services,Technologie, Meaning for everyone.

4 Applications to Health

This part of the paper divides applications of IoT technologies into the four-point health chain, extracting and processing data for the purpose of steering the health chain, safeguarding this data which can be private and Privacy, marking and object detection technologies, and finally how to evaluate the performance of a system in the Internet of Things.

4.1 Data extraction and processing in IOI

"Big Data" is gradually and invasively replacing relational databases that have existed since the 1990s. These giant databases allow for the management of more data, their disparity and heterogeneity, Real-time access to information whether distributed or centralized. With the rise of the Internet of objects, sensors embedded on mobile devices make it possible to trace more and more data (temperature, pressure, GPS position, speed, luminosity, heart rate, etc.). This situation complicates more and more the task of analysts and data mining who are faced with the problem of proposing new solutions adapted to this steadily increasing volumetry, replacing the traditional searches that are based on General on statistical models, linear regression or health [20]. Also, it should be noted that the collection, formatting and transmission of data from sensors to Cloud platforms requires a lot of energy, which can be very constraining, so that we must now consider moving tasks Of processing to devices whose energy constraints are lower than those of the sensor (Smartphone, PDA, computer). There are middleware that have been developed in this sense (MOSDEN: Mobile Sensor Data Processing Engine) [18].

On the other hand, protocols that have been used for a few years for web applications (SOAP: REST: Representational State Transfer) are also used for object exchanges.

Data extraction languages (SPARC QL, ETL: Extract Transform and Load) facilitate the extraction and processing of data from various sources, distributed or non-distributed databases, relational databases or NoSQL. These languages, standards and protocols also aim to integrate the Internet of Things with Cloud Computing to give birth to the Cloud of Things [20], in order to create more intelligent applications, to make the data collected On objects that are more accessible and, above all, more relevant and meaningful.

The applications are diverse and varied in the field of healthcare, from production to management of relationships between all stakeholders in the healthcare chain including the client. At the production level, sensors (RFID tags) are integrated on products (pallets, bags, bins) to store information about their contents, traceability information (operations performed on products, origin of the product and its components , ...). Intelligent production line applications have emerged via RFID readers mounted on conveyors to automatically route items or pallets according to the operations to be performed or their contents.

On the other hand, it is possible to integrate the production chain and the Enterprise Resource Planning (ERM) systems, so that the data transmitted By intelligent sensors in the production chain can be used for local and global decision making for product manufacturing [18]. This integration also allows the strategic and tactical management of the company at global level (planning, management, production scheduling, resource management), and serves to optimize the maintenance of equipment in all areas (electric vehicle, wind, solar , Aeronautics, transport, ...). An example of an application is that of BMW which uses Win River technology to connect its on-board computer to telecommunication operators in order to dialogue in real time with cloud services (weather, traffic) for vehicle maintenance in case of breakdown.

This service also makes it possible to retrieve information on the condition of important vehicle parts using integrated sensors, as well as statistics (information on the state of the road, flow of traffic) and More accurate diagnostics.

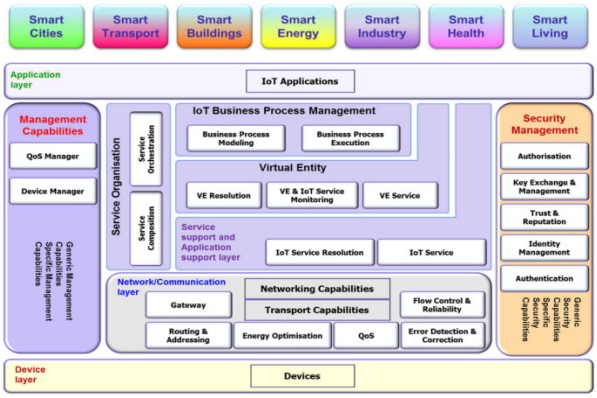


Figure 5 : Detaiiled IOT Layered Architecture(Source IERC).

4.2 Managing security, privacy, and privacy with the IoT

Security is one of the major problems of the internet in general, and Internet of the Objects in particular. More than 70% of connected objects are vulnerable to attacks according to the VDC analysis firm. The EagleEye program, developed by Dan Tentler, would take control of nearly one million webcams via the Showdan connected object engine. The security issue revolves around a few central issues, including the protection of personal and confidential data from intrusions of any kind (unauthorized reading, falsification, identity theft, espionage). There is also the protection of the routing channels of information against passive attacks (listening to the network), protection of the integrity of the sensors. Attention must therefore be paid to the protection of connected peripherals against the various forms of piracy: diversion of sensors, access to video surveillance data and other confidential data, authentication. To manage all these problems, protocols, algorithms and architectures have been proposed, among which APHA (Aggregate-Proof based Hierarchical Authentication scheme) for the secure transmission of data in ubiquitous and layered networks [15]. We also cite the algorithms

Cryptographic applications using arithmetic operations that offer relatively light architectures in terms of computation and resources, adapted to the environment of the embedded and the IoT [17]. Other protocols offer

On the structure of the web resources by coupling information related to the context of the object to the usual identification information to solve the problem of permissions and decentralized access control to the resources and information published by the objects in the Cloud.].

4.3 Marking and detection technologies in IoT

Several techniques are used in tagging and identification in the IoT. DNA markers (nitrogen base chain: guanine, cytosine, thynine, adenine) use a mixture of products (liquid or powder) to develop a unique identification and authentication code for the labeled product, typically petroleum products And pharmaceutical, and technical parts of the industry. Based on synthetic DNA, DNA markers are theoretically impossible to counterfeit, given the number of combinations on the base segments (GC-T-A) [33]. RFID allows, using an identifier contained in a chip, to address many more objects than bar codes

and OCR (Optical Character Recognition) codes. RFID tags are now widely embedded in sanitary objects (containers, wagons, pallets, cartons, trolleys, plastic containers, garments) for traceability, real-time tracking and product flow coordination. The 3D digitization of the manufacturing area of the product makes it possible to put in place a so-called hybrid code which makes it possible to identify the object in a unique way. Other techniques such as the matrix code (Datamatrix, QR code, Maxi code) [34] are used for the sorting of packages in the transport of mail. Marking techniques such as

Digital code, nano-plotters (electroluminescent, nano-particles), biometrics (facial recognition, fingerprints, voices), holograms, are also used for the identification, traceability and coordination of sanitary objects.

4.4 Performance evaluation in IoT

Measuring the performance of a system in the IoT still poses many problems because the objects must be tested under real conditions of use. Another difficulty is due to the fact that the systems in the IoT are mostly based on sensors / actuators and require a more or less strong interaction with human operators. They can integrate multiple technologies and disciplines, making it much more complex to test and evaluate the performance of such systems. There are several test solutions in different degrees of realism, classified according to their architecture (2-Thirds or 3-Thirds) and the domains covered as illustrated in Figure 5. The test benches can propose generic functionalities (MoteLab, WISBED, SenseLab) or application-specific (CitySense, Friedrichshafen, Oulu Smart City). Depending on their architectures, evaluation environments can be categorized into two categories, depending on whether the architecture is two-thirds (Sensor-Server) or three-thirds (Sensor-Server-Internet). For 2-tier architectures we can mention MIRAGE, Vinelab, City Sense, FrONTS, whose limit is their inability to support the network layer. 3-tier architectures such as TWIST, INDRIYA, take into account the network layer to facilitate communication between objects and test servers, offering more flexibility and performance gains. These test solutions can integrate one or more operating systems generally used in the embedded domain (TinyOS, iSense, MoteRunner, Contiki, Sunspot) [15].

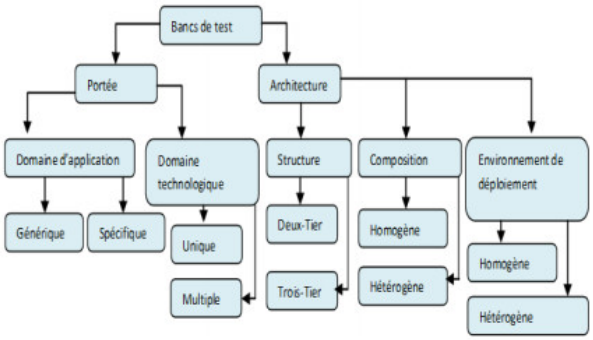


Figure 6: Classification of Assessment Technologies TheIoT

5 Conclusion

The construction of trust goes through multiple paths. The governance of the Internet Future must be multipolar and multi-actor. The current governance of the Internet is a tool.In the context of the IoT, it will have to evolve to apply to an environment.The benchmarks of private life, territorial sovereignty or

time and memory are themselves in a phase of transformation. Regulatory measures should beAnd to ensure shared reordering of values.

To "think" internet governance in political terms th ideal-type of neutrality of theTechnique is necessary, but not sufficient. The digital economy and politicsInsidious manner of powerful constraints and, at the same time, mobility and connectionOpen up unprecedented possibilities for the democratic functioning of societies andFor better shared access to knowledge. Well-informed of these paradoxes, the institutionsAt national, European and international levels,Research in the humanities and social sciences, as well asResearch and development efforts.At the end of this study, we can only note the need for further research.Without answering all the questions posed by the Internet of the Objects, we mainlyRaised new ways of posing problems whose solutions are far from beingEvident: privacy, governance, standardization, cooperation, etc.

Each of these subjects deserves to be treated as such, and in a coordinated way, in thePerspective of the development of the Internet of Things. In addition, the need for reflection in thisSector seems urgent, both the development of technical solutions and newApplications is accelerating every day.In these circumstances, it is likely that a research effort should beTo address technical, economic, social and legal issues simultaneously.It is a prerequisite for a better understanding of the Internet ofAnd the ability of Europe to defend the values of its citizens in the new environmentdigital.

REFERENCES

[1] S. Bandyopadhyay, M. Sengupta, S. Maiti, and S. Dutta, “Role of Middleware for Internet Of Things,” International Journal of Computer Science & Engineering Survey (IJCSES), vol. 2, no. 3, pp. 94–105, 2011.

[2] M. A. Chaqfeh, N. Mohamed, P. O. Box, and A. Ain, “Challenges in Middleware Solutions for the Internet of Things,” International conference on Collaboration Technologies and Systems (CTS), no. IEEE, pp. 21–26, 2012.

[3] F. Schreiber, M. Camplani, Romolo Fortunato, M. Marelli, and G. Rota, “Perla: A language and middleware architecture for data management and integration in pervasive information systems,” Software Engineering, IEEE Transactions on, vol. 38, pp. 478–496, 2012.

[4] K. Henricksen, J. Indulska, and S. McFadden, Ted Balasubramaniam, “Middleware for Distributed Context-Aware Systems,” in On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, R. Meersman and Z. Tari, Eds. Springer Berlin Heidelberg, 2005, pp. 846–863.

[5] E. Guinard, Dominique and Trifa, Vlad and Mattern, Friedemann and Wilde, “5 From the Internet of Things to the Web of Things : Resource Oriented Architecture and Best Practices,” in Architecting the Internet of Things, F. Uckelmann, Dieter and Harrison, Mark and Michahelles, Ed. Springer Berlin Heidelberg, 2011, pp. 97–129.

[6] J. Sen Debasis Bandyopadhyay, “Internet of Things - Applications and Challenges in Technology and Standardization,” Wireless Personal Communications, vol. 58, no. 1, pp. 49–69, 2011.

[7] ITU-T Study Group, "New ITU standards define the Internet of Things and provide the blueprints for its development," ITU, 2012. [Online]. Available: http://www.itu.int/ITU- 74 T/newslog/New+ITU+Standards+Define+The+Internet+Of+Things+And+Provide+The+Blu eprints+For+Its+Development.aspx.

[8] A. Bassi, "Introduction (IoT-A)," IoT-A, 2013. [Online]. Available: http://www.iota.eu/public.

[9] T. Teixeira, S. Hachem, and N. Georgantas, "Service Oriented Middleware for the Internet of Things: A Perspective)," in Towards a Service-Based Internet, vol. 257178, no. 257178, J. Abramowicz, Witold and Llorente, IgnacioM. and Surridge, Mike and Zisman, Andrea and Vayssière, Ed. Springer Berlin Heidelberg, 2011, pp. 220–229.

[10] J. Soldatos, N. Kefalakis, and N. Leontiadis, "ASPIRE Programmable Engine (APE)," ASPIRE Consortium, vol. 3, 2011.

[11] A. Salehi, M. Riahi, S. Michel, and K. Aberer, "GSN, middleware for stream world," in 10th international conference on mobile data management (MDM 2009), May, 2009, pp. 18– 20.

[12] P. Eisenhauer, Markus Rosengren and P. Antolin, Hydra: A development platform for integrating wireless devices and sensors into ambient intelligence systems. Springer, 2010, pp. 367–373. 75

[13] J. Domingues, A. Damaso, R. Nascimento, and N. Rosa, "An Energy-Aware Middleware for Integrating Wireless Sensor Networks and the Internet," International Journal of Distributed Sensor Networks, vol. 2011, pp. 1–19, 2011. [30] C. S. M. Villa, D. Rotondi, M. Comolli, (TXT), A. M. H. (Siemens) H.–P. Huth, C. K. (City), H. T. (inIT), and J. C. (EMIC), "WP 1 – PLUG&WORK IOT REQUIREMENT ASSESSMENT AND ARCHITECTURE," 2010.

[14] D. Yazar and A. Dunkels, "Efficient application integration in IP-based sensor networks," in Proceedings of the First ACM Workshop on Embedded Sensing Systems for EnergyEfficiency in Buildings, Berkeley, California: ACM, 2009, pp. 43–48.

[15] D. Guinard, I. Ion, and S. Mayer, "In Search of an Internet of Things Service Architecture : REST or WS- *? A Developers ' Perspective," in Mobile and Ubiquitous Systems: Computing, Networking, and Services, Springer, 2012, pp. 326–337.

[16] Z. Song, A. C. Alvaro, and R. Masuoka, "Semantic Middleware for the Internet of Things," 2010. [34] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no. elsevier, pp. 2787–2805, 2010.

[17] K. Paridel, E. Bainomugisha, Y. Vanrompay, Y. Berbers, and W. .Meuter, "Middleware for the internet of things, design goals and challenges," CAMPUS 2010, vol. 28, no.Electronic Communications of the EASST, pp. 1–7, 2010.

[18] Sh.zarghami, "Architectural support for dynamic homecare service provisioning," 2013.

[19] Mohammad Zarifi Eslami, “Service Tailoring: A Method and Tool for User-centric Creation of Integrated IT-based Homecare Services to Support Independent Living of Elderly,” Twente, 2013.

[20] A. H. Maya Daneva, Luigi Buglione, “Software Architects’ Experiences of Quality Requirements: What We Know and What We Do Not Know?,” Proceedings of REFSQ, pp. 1–17, 2013.

[21] R. S. Peter B. Seddon, “Towards the improved treatment of generalization of knowledge claims in IS research: drawing general conclusions from samples,” EJIS 21(1), pp. 6–21, 2012

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

The GeoRSS Model and the Design of a Ranking Algorithm in Semantic Web

Anas Bennani, Samira Khoulji, Kerkeb Mohamed Larbi
Information Systems Engineering Research Group, Dept. Telecommunications.
National School of Applied Science of Tetouan,
University Abdelmalek Essaadi, Morocco;
anasbenn@gmail.com; khouljisamira@gmail.com; kerkenml@gmail.com

ABSTRACT

E-commerce has become nowadays an influential actor in the international trade industry and has greatly contributed in the economic development of nations. The fact of dealing with this field along with the evolution of technologies renders the e-commerce more usable by internet users. The semantic web offers advanced technical solutions relying on the structuring and organization of the existing data on the web.

Throughout this Paper, I will tackle the issue of e-commerce in relation to geolocation, while at the same time offering the internet user more freedom to detect the surface of the geographical area where he/she is located by a proposing a new geographical coordinate classification algorithm.

Keywords-E-commerce, semantic web, Algorithm for Ranking, GEORSS model, geo owl,OGC,FOAF ontology, GOODRELATIONS, geographic information.

1 Introduction

E-commerce is defined as the sale or delivery of goods and services by information networks and methods specifically designed for receiving or making orders. This is the definition given by the World Trade Organization (WTO); an international body that deals with the rules of trade between nations. Internet commerce has become essential for many countries of the world. In 2016, French consumers spent 72 billion euros online, which is an increase of 14.6 percent compared to the situation in the previous year (2015). They carried out 103 billion online transactions, equivalent to 33 orders per second, growing by 23% , according to the data published recently on Thursday 26 January by the French Federation of E-commerce and Distance Selling (FEVAD).

To complete the buying / selling process online, all the procedures, from the product's exposure to its payment, have to be followed in an effective and secure way. Products have to be well described. And their announced prices and the after-sales service have to be well-explained. As long as the products are properly presented on thecompany portal,they will become clearly visible to whoever clients looking for them.

The technology which allows the structuring of data on the web and a better organization is the semantic web. In this regard, Tim Berners Lee has demonstrated his idea by stating the following: "building a web

DOI: 10.14738/tmlai.54.3322
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3322

of data that might be directly and indirectly processed by the machines so as to help their users to create new knowledge [1] ".

In order to create data in the semantic web, we mainly use an ontology [2]. It is a formulation of a concept. As to our Semantic Web area, the World Wide Web Consortium (W3C) has defined ontology as follows: «An ontology is a formal specification of a common set of terms used to depict and represent an area. It defines the terms used to describe and represent an area of knowledge". [3]

In parallel, the present paper will be concerned with the geographic information System (GIS) that has become a constituent part of our daily lives. And this paper will especially make use of the GIS in the GoodRelations ontology: e-commerce ontology. The GIS is used, as a case in point, to specify the location of a new boutique or a particular store in favor of a company or a trader. More recently, the growth and use of (GIS) were exposed by the implementation of the Web-based tools. The use of geographic information has greatly increased; however, many users will be largely oblivious to the role it plays in their activities. This is because GIS per se is considered as a fairly rare target; rather, it normally creates the scenario under which the topic of interest is remarkable. [4]

The main goal of the present paper is to come up with a new approach of using GoodRelations with the GEORSS model. The latter allows the representation of geographic information on the semantic web. The rationale behind choosing this process that I will study throughout the present paper is to offer any person or tourist the possibility to determine a distance over which he/she may have information about the stores or shops that surround him/her. I suggest that the owners or sellers have to make use of the GoodRelations ontology to present their products and, in the same vein, should make sure that their geographic coordinates are implemented via the properties dedicated to the geolocation. The method I propose here plays an indispensable role in the tourism industry, especially for the benefit of tourists who want to travel around a city without a guide.

2 Existing Ranking Methods

2.1 Spatial data model: GeoRSS

The Geographically Encoded Objects for RSS feeds (GeoRSS) is a model recommended by the World Wide Web Consortium (W3C). It represents a reference vocabulary for the description of the geo-spatial properties of Web resources. The following illustration shows the UML modelling of GeoRSS. This paper will examine the box-sizing property of the model and will evolve its functionality [5].

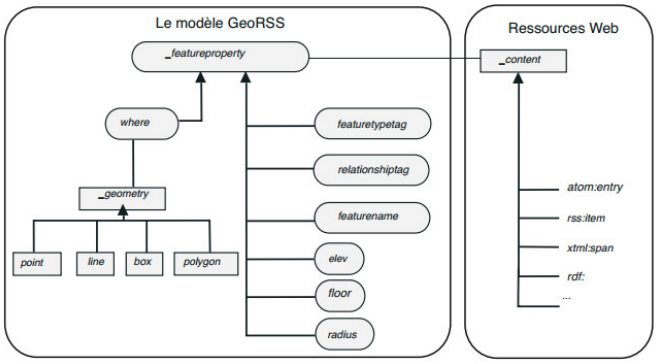


Figure 1: An Extract of the GeoRSS model

The Geography Markup Language (GML) is an improved version of the GeoRSS model. It is an XML grammar for encoding, manipulating and exchanging spatial data. This standard is developed by the Open Geospatial Consortium (OGC) to ensure the interoperability of data, particularly in the fields of Geographic Information Systems and Geomatics. Such sort of language, namely (XML), covers various types of objects which describe spatial data, including entities, coordinate reference systems, geometries, topologies, time, units of measurement and generalized elements [6]. GML is a set of XML schemas that specify an open interchange format for exchanging data and that permit the building of specific models for specialized domains, such as urban management.

Figure 11: The box-sizing property of the GeoRSS model

The following section will cover the benefits and the added value of this item to the geo-location and its impact on e-commerce.

2.2 GoodRelations Ontology For E-commerce

The GoodRelations ontology covers a comprehensive set of procedures dedicated to e-commerce, starting from the description of products until payment. This vocabulary is developed by adding extra semantic codes on Web content [7].

In the present work, I will deal with the issue of e- commerce vis-a-vis the geographic location of stores. The following illustration [7] technically shows the properties of the gr:location class.

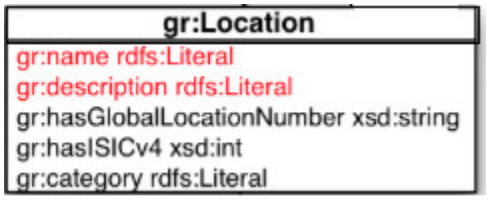


Figure 12:The GoodRelations Location Class

The objective is to suggest a new way that could allow one to choose a distance that will enable one to delimit the geographical zone where one is located. The desired distance has a direct impact on the surface zone, i.e., if the distance is vast, the surface zone will also be large. Subsequently, thanks to the GoodRelations ontology, one may recognize the various stores located in the area as well as their features and some information on the products that are for sale.

2.3 Friend of a Friend (FOAF) Ontology ForPerson

Friend Of a Friend (FOAF) is a widely used ontology in the semantic web that allows to describe people by providing all the available information on them, such as their Web pages, surnames and names, etc. Additionally, this vocabulary offers the possibility to know the relationships they have with each other[8].

In this case study, the ontology will have information about the people who may become customers. In this regard, the storeowners who are located in the delimited geographical area can use that information to capture those people's attention. As a case in point, "a person loves eating chocolate": such information is semantically recorded in the ontology code. In that, if a store, for instance, has chocolate pieces on sale, that person will be automatically notified of the offer.

3 The Spatial Datum in the Semantic Web

3.1 Determination of GPS Coordinates.

In order to give importance to the geographical location of a particular person, I suggest an approach which will improve the detection of the geographical area. To this end, we can use the GeoRSS model that we have presented in the above-mentioned chapter under the property "Box". The suggested scenario is the following:

The person who may become a customer with a store is found in a location specified by the geographic latitude and longitude coordinates (Lat1& Lon1). These parameters represent the first coordinates "lower corner" of the Box. The added value of this item is summarized as follows: so as to establish the geographic coordinates (Lat2 & Lon 2) of the "upper Corner" of the Box, one could determine the distance at which one may be surrounded by the rectangle traced by the property Box, with an orientation equal to 45 ° = Pi/4 Radians. Normally, if we are to detect the geographic coordinates of a particular destination, we will need to know the coordinates of the departure, of the distance covered and of the orientation angle. Since its parameters are known, we make use of the formula developed by Thaddeus Vincenty[9] to find those coordinates following an ellipsoidal model of the Earth.

Lat2 = ASIN(SIN(lat1)*COS(d/R) + COS(lat1)*SIN(d/R)*COS(Orient))

Lon2 = lon1 + ATAN2(COS(d/R)-SIN(lat1)*SIN(lat2) , SIN(Orient)*SIN(d/R)*COS(lat1))

Knowing that R: Ray of the Earth = 6371 Km; d = distance in Kilometers (Km); Orient = Orientation in radians.

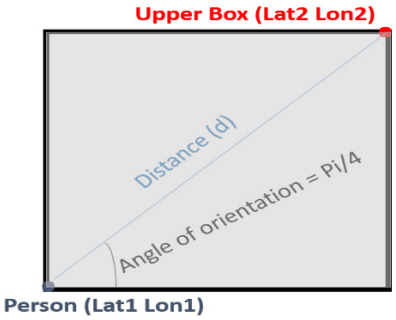


Figure 13: Setup of the Property Box

The GeoRSS -Simple model makes use of the World Geodetic System (WGS84). Geodesy, in this respect, is a coordinate system used to indicate the trajectory corresponding to the minimum distance between two reference points on a surface. As to the sphere, it is the arc of a great circle. For instance, by knowing the position of two points represented in A and B on a sphere, we could, therefore, so as to estimate the distance between them, calculate the curvilinear abscissa S (AB) on the great circle passing through A and B, as it is shown in the following figure. Geodesy is the science that studies the Earth's geometric shape, dimensions and gravity field. Its main objective is to develop terrestrial reference systems to which any georeferenced data user or creator can have access through the intermediary network.

The Adoption and implementation of such reference systems constitute an essential standardization tool for the geographic information in particular and the positioning in general. Nonetheless, the calculations of a spherical model are accurate to within 0.3%. Vincenty's formula is accurate to within 0.5 mm or 0.000015'on the ellipsoid used. Note that the accuracy of theWGS-84 model will be lower according to the Earth's position, but generally in the range of 0.003% [10].

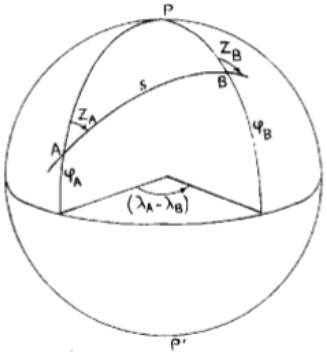


Figure 14: The Position of Two Points on a "Geodesic" Sphere

3.2 Algorithm Conditions for the Ranking of Geographic Datum

The algorithm that is proposed for this item requires conditions at the organizational level. Admittedly, the author of this paper had a Geographic Coordinate Database which included the stores- in ascending order- of the region or the city where he was located. This starts with the latitude and, then, longitude coordinates.

This paper considers that those coordinates are gathered in a semantic way thanks to the GoodRelations ontology and are sorted out in ascending order and stored in a table. Here is an example:

Table 13: An Example of sorting out the geographic coordinates of stores

	LATITUDE	LONGITUDE
STORE 1	0,7673214	0,0236234
STORE 2	0,7673300	0,0236261
STORE 3	0,7673416	0,0236125
STORE 4	0,7673675	0,0236259
STORE 5	0,7673780	0,0236395
STORE 6	0,7673865	0,0237264
STORE 7	0,7673868	0,0227456
……..		

As far as the programming language is concerned, the coordinates of the store 4 will be stored in a table as follows:

Latitude = Tab [3] [0] = 0.7673675 and Longitude = Tab [3] [1] = 0.0236259.

3.3 Algorithm

```
Variables
i, j as Integer number (* to increment*)
NbreStrore as Integer Number(* Number of Stores*)
lat1, lat2, lon1, lon2 as real number
Table tab()()(* Two-dimensional array declaration
containing geographic coordinates*)
TableCGC()()as real number(* Coordinate Geographic
candidate table declaration which
will contain the coordinates of the
stores located in the box *)
Begin
  Write"The number of stores detected during the search"
Read (NbreStore)
Resize tab(Nbrestore-1)(NbreStore-1)
ResizeCGC(Nbrestore-1)(NbreStore-1)
CGC(0)(0) ← 999(*Initialization of Latitude
dimension *)
CGC(0)(0) ← 999(*Initialization of Longitude
dimension *)
j← 0 (* To increment CGC table *)
      ForiFrom 0 ToNbreStore-1
```

```
If (( tab(i)(0) <= lat2 )
and(tab(i)(0) >= lat1 )
and(tab(i)(1) <= lon2 )
and(tab(i)(1) >= lon1 ))
CGC(j)(0) = tab(i)(0);
CGC(j)(1) = tab(i)(1);
Nextj
End If
Next i
End For
Read (CGC)
End
```

3.4 The Positioning of stores in relation to a particular person

Once the geographic coordinates of the listed stores pertain to the surface of the Box via the algorithm, we could recognize the distance between the store and the position of the person requesting the information, as well as the orientation (Angle to the North). And this is thanks to the following formulae:

d = ACOS(SIN(lat1)*SIN(lat2)+COS(lat1)*COS(lat2)*COS(lon2-lon1))*R

Orientation = ATAN 2(COS(lat1)*SIN(lat2)-SIN(lat1)*COS(lat2)*COS(lon2-lon1), SIN(lon2lon1)*COS(lat2))

Here follows a mathematically equivalent formula which is, however, less prone to errors of a rounding nature as far the short distances are concerned:

d=2*asin(sqrt((sin((lat1-lat2)/2))^2 + cos(lat1)*cos(lat2)*(sin((lon1- lon2)/2))^2))

Such formulas are set for a spherical model of the Earth and constitute, as a result, the approximations. [11]

4 Conclusion

In this research paper, I have, on the one hand, shown interest in e-commerce and its link with the spatial dimension. The goal, as mentioned earlier, is to offer internet users, generally, the opportunity to conduct the desired search on the geographical area where they are located and to enable them to obtain as much information as possible.

The fact of combining e-commerce with geolocation in a structured manner contributes to a better organization of the data. I have also demonstrated, on the other hand, that e-commerce has become a highly intelligent sector in the sense that it gives all Internet users, who are now attached to technology more than ever before, the possibility to automatically carry out any sophisticated commercial operation with the minimum number of interventions.

The result mentioned in this paper will contribute to the development of the tourist and economic sectors of the countries because it allows people in general and tourists in particular to identify the shops around

them and know the types of their products offered for salefrom research. This paper also facilitates the sales / buying process either at the seller level by targeting buyers who are interested in their products, or aside from the buyer who finds the desired products that he will be able to buy them and at the minimum of time.

REFERENCES

[1] Tim Berners-Lee, Book: “Weaving The Web”, 1999

[2] Grigoris Antoniou, Frank Van Harmelen, “A Semantic Web Primer”, Fltebook,The Mit Press Cambridge Massachusetts, London, England 2004.

[3] Liyang Yu, “A Developer Guide To The Semantic Web”, Springer, Isbn : 9783642159695,Pp. 136,137, 2011.

[4] Glen Hart,Catherine Dolbear, Ebook “Linked Data A Geographic Perspective”, Isbn: 13: 978-1-4398-6997-0, 2013.

[5] Wafa Mefteh, Thesis : “An Ontological Approach For Modeling And Reasoning On Trajectories: Taking Into Account The Thematic, Temporal And Spatial Aspects ”, The University – La Rochelle -2013

[6] G. Percivall, “Opengis Reference Model”, Document Number Ogc 03-040,Version 0.1.3. Open Gis Consortium, Inc., 2003.

[7] Martin Hepp, Official Web Site Http://Www.Heppnetz.De/Projects/Goodrelations/

[8] Official Web Site Of Foaf Ontology Http://Www.Foaf-Project.Org

[9] T. Vinccnty, “Direct And Inverse Solutions Of Geodesics On The Ellipsoid Wlth Application Of Ested Equations”, Directorate Of Overseas Surveys Of The Ministry Of Overseas Developmen, 1975.

[10] Official Web Site Of National Institute Of Geographic And Forestry Information http://Geodesie.Ign.Fr

[11] H. P. Garnir, Institute Of Nuclear, Atomic And Spectroscopic Physics, “Calculation Of The Distance Between Two Points”, University Of Liège Belgium, Training Day At The University Of Liège, 2004

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Gamification for Arabic Natural Language Processing: Ideas into Practice

Imad Zeroual, Anoual El Kah and Abdelhak Lakhouaja
Computer Sciences Laboratory, Faculty of Sciences, Mohammed First University
Oujda, Morocco
mr.imadine@gmail.com, elkah.anoual.mri@gmail.com, abdel.lakh@gmail.com

ABSTRACT

Gamification is a novel IT development that focuses on how to address the use of games and apply game design techniques in a non-gaming context. Recently, it has significantly attracted several researchers' attention concerned with different fields such as education, business, and information retrieval, among other fields. Most Natural Language Processing (NLP) applications use large accurately labelled data sets to achieve good performance, but these data sets are hard to obtain. The standard method to produce labelled data has been through manual labour that demands a strong engagement and a heavy financial participation. For these reasons, gamification may offer smooth possibilities for wide improvements in this field. However, gamification by itself may not yield the widely-expected results if the incentives are not strong motivators. Further, based on a previous experience, we strongly believe that even financial incentives are not enough for the participants to increase their contributions. Therefore, we could incentivise the participants, through gamification, and through their need for language acquisition. Thus, language-based games are adopted in this project as a direct incentive and a stronger engagement to ensure the benefits for both NLP practitioners and the language learners. In this paper, we present the first Arabic project that targets all the means to promote and explore the possibilities for research and practical applications of using games and gamification for Arabic NLP. Here, we shed light over the advantages of using a gamified approach in corpus annotation, named entity recognition, and word sense disambiguation.

Keywords—Gamification, Games design elements, Natural language processing, language-based games, Arabic language.

1 Introduction

Basically, gamification originates from the computer games industry. It is based on applying game thinking and game mechanics or game design techniques to non-game activities in order to engage users, solve problems, and drive behaviour [1]. Gamification is considered to be an effective approach to increase users' motivations in non-game activities or towards using technology; thereby, increasing the quantity and quality of the output of the given activities [2].

Since 2015, millennials became the largest generation in the workforce [3]; yet, the world's population spends roughly 3 billion hours a week playing video games [4]. Thus, gamification is becoming increasingly

DOI: 10.14738/tmlai.54.3323
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3323

important especially with the tremendous growth of the smartphone and the increased use of the Internet. This raises the gamification implementations in education [5], information retrieval [6], healthcare [7], business [8], tourism [9], and insurance markets [10], among other fields.

On the other hand, most Natural Language Processing (NLP) applications require a strong engagement and greater participation which is often not available to reach targeted purposes or solve some issues, primarily for morphologically complex and resource-poor languages such as Arabic. Due to these convincing reasons, we believe that gamification and language-based games could incentivise the learners to continue for their contributions and ensure the benefits for the NLP practitioners. For example, games can be used to collect large numbers of annotations provided by sufficient numbers of learners who are motivated to play. Thereby, the aggregated annotations can be used as labels that complement a previous automatic annotation and may replace the post-editing by a domain expert, which is often difficult to have someone available, at least for the Arabic language.

In this paper, we present the first Arabic project that aims promoting and exploring the possibilities for research and practical applications of using gamification for Arabic NLP. The main objective is to provide recommendations collected from existing research and to discuss how the Arabic NLP researchers and practitioners can benefit from using games and gamification. Further, a preliminary approach for developing and testing some proposed language-based games is presented. This gamified approach is primarily used in corpus annotation, named entity recognition, and word sense disambiguation.

The paper is arranged in five main sections focusing on the basic aspects of gamification and the possibilities of using it for Arabic NLP. In section 2, a background information regarding Arabic NLP is provided. In section 3, we briefly describe the standard basics of gamification and the known outcomes of its implementation. Further, guidelines and recommendations for developing games primarily language-based games are presented. In section 4, the proposed gamified approach and a set of preliminary games are described. Finally, we conclude this paper in section 5.

2 Arabic Natural Language Processing

Arabic is currently the most used Semitic language [11]. It is expanding in the world in an area extending from the Arabian/Persian Gulf in the East to the Atlantic Ocean in the West. According to UNESCO, the Arabic language is spoken by more than 422 million native speakers [12] around 29 countries and among 1.6 billion Muslims worldwide use it to perform their daily prayers [13]. Moreover, the presence of the Arabic language on the internet grew around 6.091% in the last fifteen years (2000–2015), it is the highest growth of the ten top online languages, and it is the fourth most spoken language on the internet [14]. Adding the fact that teaching Arabic as a foreign language has become a global educational enterprise [15].

2.1 Arabic language features

The standardization of Arabic lexicon and grammar are deeply rooted and well established a long time ago. The morphology of Arabic differs in the structure of affixes from Indo-European languages [16] (e.g. English, German, French, Hindi). Arabic is a highly productive both derivationally and inflectionally [17]. These two complex paradigms are based on the interaction between roots and patterns which have intrigued lexicographers and morphologists for centuries.

Basically, Arabic language consists of three main categories [18], [19]: Noun "اسم" <Asm>, Verb "فعل" <fEl> and Particle "حرف" <Hrf>. In addition, each one of these categories has dozens of subcategories [20]. One of the most challenging and interesting property of Arabic is the concatenative morphemes, where one word could represent a whole sentence through sequential concatenation. For example: the Arabic word "أَسَأَلْتُمُونِيهَا" <AasaAalotumuwniyhaA> means in English "Do you have already asked me for it".

The omission of diacritics (short vowels) in written Arabic is another source of some difficulties in several automatic processing systems of Arabic language [21]. When these vowels are omitted it is left for the reader to infer; knowing that the vowels can encode grammatical category or feature information. Typically, a root or a lexical item consists of three consonants. The vowels add grammatical information when attached to these consonants. For example, the three consonants ktb can stand for the verb "كَتَبَ" <kataba> meaning "he wrote", or for the plural noun "كُتُبٌ" <kutub> "books".

Another characteristic of the Arabic language is the free word order nature in the sentence which makes parsing Arabic sentences one of the most difficult tasks. i.e., we could easily change the order between the subject and the verb; yet, it is not necessary that they must agree with one another in number (singular or plural) for all cases. For example:

1. The boy hit the ball

الْوَلَدُ ضَرَبَ الْكُرَةَ or ضَرَبَ الْوَلَدُ الْكُرَةَ

2. The children watch television

الأَطْفَالُ يُشَاهِدُونَ التِّلْفَازَ or يُشَاهِدُ الأَطْفَالُ التِّلْفَازَ

Although the free order of the subject "الْوَلَدُ" <Alowaladu> "The boy" and the verb "ضَرَبَ" <Daraba> "hit"; yet, the subject–verb number agreement of the subject "الأَطْفَالُ" <AloATofaAlu> "The Children" and the verb "يُشَاهِدُ/ يُشَاهِدُونَ" <yu$aAhidu/yu$aAhiduwna> "watches/watch" has changed, both sentences are grammatically correct and have the same meaning. And this becomes more complex in longer sentences.

2.2 Arabic text processing

In this section, we present some well-known tasks in Arabic NLP fields and which will be under consideration for the gamification implementation.

2.2.1 Corpus annotation

One of the main aims being addressed in corpus linguistics is to develop new forms of annotation and improve the accuracy of automatic annotation. The following tasks are some of the key features of text annotation:

- Lemmatization: It is relating words written in different syntactic forms to their canonical base representation (i.e., lemma). The lemma is a dictionary lookup form which can relate different word forms that have the same meaning [22]. For instance, the lemma for the verbs "يَكْتُبونَ" <yakotubwna> "they write" and "تَكْتُبُ" <takotubu> "she writes" is "كَتَبَ <kataba> "to write".

- Part of speech tagging: It is a basic task in many fields primarily corpus linguistics and NLP. It aims to automatically assign a morpho-syntactical label (e.g., noun, verb, and adjective) to every word in a written text. Further, this task allows simple syntactic searches to be performed.
- Parsing: It is the natural successor to part of speech tagging. Basically, it provides a dependency tree as an output. The goal of parsing is to predict for each sentence or clause an abstract representation of the grammatical entities and liaisons between them. Consequently, the parser assigns a fully labelled syntactic tree or bracketing of constituents to sentences of the corpus [23]. For example, Figure 1 exhibits a parse Tree of the first verse of chapter 113 (surah Al-Falaq) of the Holy Quran "قُلْ أَعُوذُ بِرَبِّ الْفَلَقِ" "Say, 'I seek refuge in the Lord of daybreak'" which is parsed using NLTK [24].

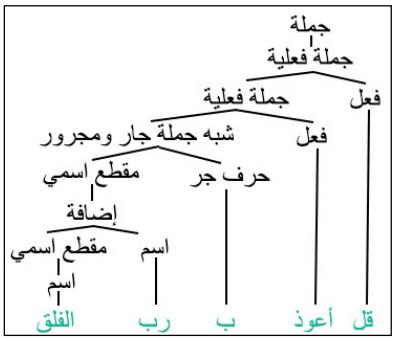


Figure 1 A parse Tree of the first verse of chapter 113 of the Holy Quran.

2.2.2 Named entity recognition

Named entity recognition (NER) is an information extraction technique that aims to locate and categorize proper nouns (e.g., names of persons, organizations, and locations) in a text [25]. NER plays a vital role in several NLP applications; it provides key information that are very useful for Question Answering and Machine Translation.

Work related to Arabic NER is rather limited and under-studied as a recent survey reported [26]. One of the main reasons that led to this situation is the absence of a large labelled corpus which is required to train NER models. In addition, the mentioned characteristics of the Arabic language make this task very challenging. Figure 2 shows a sample NER output for an Arabic sentence:

"يوم الجمعة الماضية، قام الملك محمد السادس بافتتاح مركز الدراسات والبحوث الانسانية والاجتماعية بمدينة وجدة."

"Last Friday, King Mohammed VI inaugurated the Center for Human and Social Studies and Researches in Oujda City."

يوم <DATE>الجمعة</DATE> الماضية، قام الملك <PER>محمد السادس</PER> بافتتاح <ORG>مركز الدراسات والبحوث الانسانية والاجتماعية</ORG> بمدينة <LOC>وجدة</LOC>.

Figure 2 A sample NER output for an Arabic sentence.

2.2.3 Word sense disambiguation

Word Sense Disambiguation (WSD) is a technique to determine automatically the correct or the appropriate sense of each ambiguous word regarding its context [27]. WSD is one of the most difficult

problems in NLP that negatively affects the accuracy of information retrieval systems due to the sparse nature of the queries involved. It exerts a similar influence upon machine translation [28]. It is ubiquitous across all languages but it has greater challenges in the Arabic language [29] primarily due to these two issues:

- A high level of ambiguity emerges when diacritics are omitted, especially in most Modern Standard Arabic (MSA) texts. For example, the unvowelled isolated word "كتب" <ktb> can mean the word "كَتَبَ" <kataba> "to write", "كُتِبَ" <kutiba> "written", and the word "كُتُبَ" <kutub> "books", among other meanings.
- Part of speech (PoS) category of a word reflects its syntactic and semantic role in a given context. What's more, it is estimated that the average number of possible PoS tags for a word in most languages is 2.3, whereas in MSA is 19.2 [30]. For instance, the same vowlled word "أَحْمَدُ" <AHomadu> has different meaning according to its PoS category as in the sentence "أَحْمَدُ اللَّه" "I thank Allah" it is a verb and as in "أَحْمَدُ بنُ عَبْدِ اللَّه" "Ahmed is Abdullah son's" it is a proper noun.

3 Gamification

Gamification is the use of game design elements in non-game contexts to increase users' motivations towards given activities [31]. Gamification was first mentioned in 2003 and became widely used in literature in 2010 [32]. Over these years, gamification has attracted a great deal of attention, from all shapes and sizes of research groups in academia and industry alike, in order to revolutionize the way they interact with targeted users and customers. Further, gamified approaches become more competitive and more profitable for data collection. It addresses the limitations of other methods such as the manual, automated, and crowdsourcing approaches [33]. In its core, gamified approaches are based on the same strategy as crowdsourcing. However, in this latter, the participants receive money to increase motivation; whereas, gamification incentivises them by getting entertained.

The major keys benefits of gamification are a better learning experience, instant feedback, and prompting behavioural change. Further, gamification techniques can provide comprehensive navigation and frequent indications to the user, which keep the user informed of what to do and explaining how answers are correct or incorrect [34]. In addition, in the areas of education and business, gamification can be used to fulfil most learning needs including induction and onboarding, product sales, customer support, soft skills, awareness creation, and compliance. Using wise strategy and high levels of engagement, gamification leads to an increase in recall and retention. It can drive strong behavioural change, especially when combined with the scientific principles of repeated retrieval and spaced repetition.

Several studies have integrated gamification concepts and benefits in the NLP and language learning basic skills. A previous study was conducted to design and develop an accessible web application, named Playlingua, to improve the Spanish language learning process and lecture comprehension through gamification [35]. Another study aimed to build gold standard data for word sense disambiguation using a "Game with a Purpose" (GWAP) called Wordrobe, This game consists of a large set of multiple-choice questions on word senses generated from the Groningen Meaning Bank [36].

The coming section is devoted to providing guidelines and recommendations for developing games and language-based games.

3.1 Games and their elements in Gamification

Gamification is about more than just playing games. It is defined as the concept of applying game-design thinking to non-game applications. It has been identified as a promising technique to improve students' engagement and encourage desired behaviours which could have a positive impact on learning in general and language learning in particular.

Games play a vital role in enhancing the children's knowledge and attitudes. It presents an opportunity to acquire knowledge and expose to both failure and success experiences, accordingly, the child gains good morals and attitudes [37]. During the game experience, children are motivated and interact in multi-sensory ways while enhancing their skills and renewing their motivation in every replay [38]. Therefore, more interest is given to edutainment and pedagogical-based games which bind knowledge with entertainment.

Viewing that designing games is a very multiplex activity; it is recommended that game design should be a team process. For this reason, game development teams should include interfaces specialists, game designers, as well as subject matter experts, cognitive task analysis, and game research in case consultation is needed [39].

Many effective techniques and procedures for designing games are set for the reason to design effective games that reliably reach intended objectives, as it is mentioned below [40]:

- ✓ Interface design is preferable to be colourful to grab the children's attention in the fact that children lose quickly their interest primarily toward educational tasks.
- ✓ An instant feedback after every activity or action should be provided for the reason to inform the child whether the answer was correct or needs another try.
- ✓ Wisely, clear sounds effects and instructions with suitable speed should be used to enrich the interactive experience of children.
- ✓ Icon should be accompanied with a label in order to allow children, regardless of their cognitive disability, to understand what expected from them when clicking on that icon.
- ✓ The set of games should be mapped in an understandable structure, especially if it contains multiple levels.
- ✓ The physical effort should be deduced, in a way that the game can be used efficiently and comfortably by either normal children or those with fine motor skills. This aim is fulfilled by using easy drag and drop functionalities and easy target selection facility.

According to pioneer studies [31], [41], [42] in the gamification field, the success of gamified systems will increase if the following game elements are considered in the design as much as possible:

- ✓ *Self-representation with avatar or virtual guidance;*
- ✓ *Points, scores, if they are exchanged for badges or rewards, they can become more effective;*
- ✓ *Leaderboards;*
- ✓ *Feedback;*
- ✓ *Classified Levels; to unlock new levels by finishing previous ones increases the self-confidence of the player;*
- ✓ *Story/Narrative context;*
- ✓ *Exercises (Challenges, Discoveries etc.);*

- ✓ *Time pressure (Countdown, Speed etc.);*

In addition to the previous elements, Morschheuser et al. [2] recently investigates overall instructions proposed by 25 experts from 16 different countries that have real-world gamification project experience. He claims that identifying the problems that should be addressed via gamification and clarifying the project's objectives and how they can be measured is the main criterion in gamified applications and systems.

3.2 Language-based games

There is an increasing use of games in learning contexts. They come in different technologies and are implemented in diversified educational settings. Well-designed games might have great potential for optimizing the learning achievements of the children, additionally, they can enhance their learning skills independently on time or places where they are, and while travelling, shopping or simply when being at home. The use of games for language learning is labeled game-based language learning, these games are often associated with the approach of task-based language teaching focusing on meaningful language use rather than drill-and-practice exercises [43]. Edutainment games, primarily language-based games, should take into consideration elements of engagement, motivation, and enjoyment. These games must be intuitive and use less cognitive load in order to ensure that the learning experience is essentially more engaging and worthwhile.

Despite the growing popularity and the promising claims that have arisen, there seems to be a major lack of empirical research and evidence on the added value of game-based learning [43], whilst, other several studies explored gaming for a specific purpose in language learning, such as listening strategies, and English as a foreign language, writing and speaking performance using a multimedia web annotation system [44].

The great challenge of developing educational games is to keep a balance between learning and gaming, and between challenge and users' abilities [45]; besides, integrating the motivating aspects of games with a good instructional design is critical. Nevertheless, effective techniques and procedures for designing games that reliably reach intended objectives are still not enough and do not fit a wide range of users including those with disabilities.

Unfortunately, the learning resources for Arab children, primarily Edutainment games, are very limited. Furthermore, the developed games and apps are not compatible with the culture, language, and background of the Arab children. In fact, less than 1% of available apps are Arabic-based [46]. Consequently, this project contributes to increase the Arabic resources by developing a suitable set of games for Arab children; yet, serving the Arabic NLP.

4 The proposed gamified approch and games

The main motivations of the proposed gamified approach are amusement, getting "knowledge", and learning a language. Thereby, learners will use the set of games for fun and learning purposes. On the other hand, the act of collecting their responses in the games is essential since it provides valuable contributing data for Arabic NLP and keeps traces regarding learner's status and comprehension of the game [39].

4.1 Gamified approach

The current work is a collection of games with purposes; each game targets a specific task of Arabic NLP. The proposed gamified approach is primarily used for corpus annotation, named entity recognition, and word sense disambiguation. The proposed set of games attempts to respect the recommended techniques and procedures for designing games in order to reach the intended objectives of our approach. It provides an instant feedback, a simplified pedagogical content, ensured engagement, well-designed interfaces, and last but not least, interaction that grabs the learners' attention and increases their motivation towards language learning.

The developed games are exploited to collect valuable data from the player's interactions. For example, games can collect large numbers of annotations information provided, since sufficient numbers of players are motivated to play. After the aggregation of the annotations, they can be used to replace or compliment the effort of NLP practitioners, primarily expert annotators. The games are designed for non-expert users, who will use their intuitions about language to annotate linguistic phenomena without being disturbed by technical terminology.

However, intensive research and studies should investigate the criteria followed to obtain reliable and accurate results such as:

- Participant's number: investigate which number of the participants is sufficient to affirm that the annotation is correct;
- Learners' score: discuss the achieved score which ensures that the player is able to annotate the raw data;

Level difficulty: evaluate the impact of the level difficulty on the performance of the player in the annotating results.

Besides, an evaluation of the games should be led to evaluate both the games' content difficulty and the player's performance.

4.2 Technical aspect

Recently, an emerging technology has led to focus more on cross-platform applications instead of device dependent application. This trending concept facilitates the access independently of the type of the device used. Multi-platform is a term for computing methods and programs that are inter-operating on different platforms and devices. Regarding the cross-platform advantages, we intentionally adopted the concept of building once and publishing everywhere.

The games are developed using a 2D game engine based on HTML5 named Construct2[1]. It creates games using a visual editor and a behaviour-based logic system. The exportation from this editor to most major platforms is allowed and the access from different devices is assured through its supported platforms: Android, Windows Phone, Windows Desktop, Mac Desktop, Linux Desktop…

[1] www.scirra.com

4.3 The proposed games

Basically, the following set of games contains three principal parts, each part targets respectively, PoS tagging, lemmatization, and parsing. These parts are the key features of text annotation.

The entire games share to some extent the same strategy: multiple-choices which could be dragged and dropped to the right place. This method provides less kinaesthetic efforts which make the games more suitable for children with fine motor skills.

Failure or success sound effects are involved in the games in addition to the instructions which are recorded and recited by a virtual companion. We adopt the auditor strategy for the sake of reducing on-screen text which could be distracting for some users with reading deficits. The oral instruction may enhance the listening comprehension of the player.

Players are restricted by timing that varies according to the game level (From 20 to 60 seconds); the harder the level is the more time is given. Players are rewarded with golden stars collected all over the set of games.

As mentioned, the set of games is beneficial for both NLP practitioners and the language learners. On one hand, the games are edutainment games that address learning Arabic language, on the other hand, we, as NLP practitioners, will benefit from the collected annotated data of the players' interactions. In the following, we describe the design of each game.

4.3.1 PoS tagging games

PoS tagging games consist of displaying a sentence divided into words. The player should define the correct PoS tag for each word in the given context by dragging the selected word to its correct morpho-syntactical label. **Error! Reference source not found.** presents an example of a screenshot of the game that focuses on nouns, verbs, and particles. It includes more categories in further levels such as adjectives and adverbs.

Figure 2 Screenshot of an example of PoS tagging games

4.3.2 Lemmatization games

Lemmatization games consist of binding words with their canonical base representation (i.e., lemma). **Error! Reference source not found.** demonstrates one of the lemmatization games; it is based on relating different syntactic forms to their lemma. The number of the suggested lemma in this game is restricted to four lemmas, it increases accordingly with which level of the game is played.

Figure 4 Screenshot of lemmatization game.

4.3.3 Parsing games

Parsing games aim to predict for a proposed sentence an abstract representation of the grammatical entities and liaisons between them. During these games, the player should assign syntactic labels to the component of the suggested sentence. **Error! Reference source not found.** shows a concrete example of this games' category.

Figure 5 Screenshot of a parsing game.

5 Conclusion and perspectives

The chief purpose of this project is using gamification for a relatively resource-poor language such as Arabic. In this paper, we present Arabic language features which differ from Indo-European languages. Further, we suggest the use of gamified approaches as competitive and alternative to produce labelled data for the Arabic NLP applications. Yet, they are more profitable for data collection and address the limitations of other methods such as the manual, automated, and crowdsourcing approaches.

Gamification improves participants' engagement by making their contribution to the task more efficient, the collection of data much faster, and the cost does not scale with the amount of data. As a case study, we developed a set of preliminary games. Although it is not expected for these games to match the quality of Arabic NLP experts, they are able to address the limitations of the other approaches and still provide good quality results. However, after the aggregation of the collected data, they can be used to replace or compliment the effort of NLP practitioners, primarily expert annotators.

Finally, this work is another step to demonstrate the advantages of using a gamified approach over other methods of data collection for Arabic language and to minimize the hand-correction. Later, intensive

research and studies should investigate the criteria followed to obtain reliable and accurate results such as participant's number, learners' score, and level difficulty. Besides, an evaluation of the games should be led to evaluate both the games' content difficulty and the player's performance.

REFERENCES

[1] B. Kim, *Understanding gamification*. American Library Association, 2016.

[2] B. Morschheuser *et al.*, "How to gamify? Development of a method for gamification," in *Proceedings of the 50th Annual Hawaii International Conference on System Sciences (HICSS)*, 2017, pp. 4–7.

[3] L. Naik, "New liquid workforce your competitive advantage: feature-workplace planning," *HR Future*, vol. 2016, no. Oct 2016, pp. 26–27, 2016.

[4] J. Young, "Can Library Research Be Fun? Using Games for Information Literacy Instruction in Higher Education," *Ga. Libr. Q.*, vol. 53, no. 3, p. 7, 2016.

[5] B. Taspinar, W. Schmidt, and H. Schuhbauer, "Gamification in Education: A Board Game Approach to Knowledge Acquisition," *Procedia Comput. Sci.*, vol. 99, pp. 101–116, 2016.

[6] C. I. Muntean and F. M. Nardini, "Gamification in Information Retrieval: State of the Art, Challenges and Opportunities.," in *IIR*, 2015.

[7] [7] D. King, F. Greaves, C. Exeter, and A. Darzi, *"Gamification": Influencing health behaviours with games*. SAGE Publications, 2013.

[8] D. Gears and K. Braun, "Gamification in business: Designing motivating solutions to problem situations," in *Proceedings of the CHI 2013 Gamification Workshop*, 2013.

[9] F. Xu, J. Weber, and D. Buhalis, "Gamification in tourism," in *Information and Communication Technologies in Tourism 2014*, Springer, 2013, pp. 525–537.

[10] S. D. Ali, "Game Theory & Gamification Applications to Property & Casualty Ratemaking," 2016.

[11] M. N. Al-Kabi, S. A. Kazakzeh, B. M. A. Ata, S. A. Al-Rababah, and I. M. Alsmadi, "A novel root based Arabic stemmer," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 27, no. 2, pp. 94–103, 2015.

[12] I. Bokova, *World Arabic Language Day*. 2012.

[13] M. B. Yassein and Y. A. Wahsheh, "HQTP v. 2: Holy Quran Transfer Protocol version 2," in *Computer Science and Information Technology (CSIT), 2016 7th International Conference on*, 2016, pp. 1–5.

[14] S. Hadžiahmetović Jurida, M. Džanić, T. Pavlović, A. Jahić, and J. Hanić, "Netspeak: Linguistic Properties and Aspects of Online Communication in Postponed Time," *J. Foreign Lang. Teach. Appl. Linguist.*, vol. 3, no. 1, pp. 1–19, 2016.

[15] M. L. Sakho, "Teaching Arabic as a Second Language in International School in Dubai A case study exploring new perspectives in learning materials design and development," British University in Dubai, 2012.

[16] I. K. Gharaibeh and N. K. Gharaibeh, "Towards Arabic noun phrase extractor (ANPE) using information retrieval techniques," *Softw. Eng.*, vol. 2, no. 2, pp. 36–42, 2012.

[17] N. Alsaedi, P. Burnap, and O. F. Rana, "Sensing real-world events using Arabic Twitter posts," 2016.

[18] A. Al-Dahdah, "The grammar of the Arabic language in tables and lists," *Beirut Maktabat Lebnan Arab.*, 1989.

[19] M. I. M. S. Ghalayini, *Jami'al-durus al-'arabiyah*. Turath For Solutions, 2013.

[20] I. Zeroual, A. Lakhouaja, and R. Belahbib, "Towards a standard Part of Speech tagset for the Arabic language," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 2, pp. 174–181, 2017.

[21] A. Chennoufi and A. Mazroui, "Impact of morphological analysis and a large training corpus on the performances of Arabic diacritization," *Int. J. Speech Technol.*, vol. 19, no. 2, pp. 269–280, 2016.

[22] M. Attia and J. Van Genabith, "A jellyfish dictionary for Arabic," in *Electronic lexicography in the 21st century: thinking outside the paper: proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*, 2013, pp. 195–212.

[23] R. Tsarfaty, D. Seddah, S. Kübler, and J. Nivre, "Parsing morphologically rich languages: Introduction to the special issue," *Comput. Linguist.*, vol. 39, no. 1, pp. 15–22, 2013.

[24] M. Shatnawi and B. Belkhouche, "Parse Trees of Arabic Sentences Using the Natural Language Toolkit," *Coll. IT UAE Univ. Al Ain*, 2012.

[25] B. Mohit, "Named entity recognition," in *Natural Language Processing of Semitic Languages*, Springer, 2014, pp. 221–245.

[26] A. Dandashi, J. Al Jaam, and S. Foufou, "Arabic Named Entity Recognition—A Survey and Analysis," in *Intelligent Interactive Multimedia Systems and Services 2016*, Springer, 2016, pp. 83–96.

[27] A. R. Pal and D. Saha, "Word sense disambiguation: A survey," *ArXiv Prepr. ArXiv150801346*, 2015.

[28] A. Zouaghi, L. Merhbene, and M. Zrigui, "A hybrid approach for arabic word sense disambiguation," *Int. J. Comput. Process. Lang.*, vol. 24, no. 02, pp. 133–151, 2012.

[29] M. Alian, A. Awajan, and A. Al-Kouz, "Arabic Word Sense Disambiguation Using Wikipedia," *Int. J. Comput. Inf. Sci.*, vol. 12, no. 1, p. 61, 2016.

[30] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Trans. Asian Lang. Inf. Process. TALIP*, vol. 8, no. 4, p. 14, 2009.

[31] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining gamification," in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, 2011, pp. 9–15.

[32] L. Lund and P. O'Regan, *Gamifying Natural Language Acquisition: A quantitative study on Swedish antonyms while examining the effects of consensus driven rewards*. 2016.

[33] T. J. Tiam-Lee and S. See, "Building a sentiment corpus using a gamified framework," in *Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2014 International Conference on*, 2014, pp. 1–8.

[34] D. Barisas, A. Duracz, and W. Taha, "DSLs Should be Online Applications," *2014 Jt. Int. Conf. Eng. Educ. Int. Conf. Inf. Technol. 2-6 June 2014 Riga Latv.*, pp. 314–319, 2014.

[35] P. Torró, E. and Lloret, E. and Moreda, "Playlingua: Exploring the Application of Natural Language Processing Technologies and Gamification in Language Learning," *Edulearn15*, no. April, 2015.

[36] N. J. Venhuizen, V. Basile, K. Evang, and J. Bos, "Gamification for Word Sense Labeling," *Proc. 10th Int. Conf. Comput. Semant.*, pp. 397–403, 2013.

[37] R. Skiada, E. Soroniati, A. Gardeli, and D. Zissis, "EasyLexia: A Mobile Application for Children with Learning Difficulties," *Procedia Comput. Sci.*, vol. 27, pp. 218–228, 2014.

[38] C. Foscari and F. Caon, *LE PLAISIR DANS L'APPRENTISSAGE DES LANGUES Un défi méthodologique*. Perugia.

[39] B. Review, "Handbook of Research on Educational Communications and Technology," *Educ. Technol.*, vol. 13, pp. 260–263, 2010.

[40] A. Dhir, "Examining the Educational User Interface , Technology and Pedagogy for Arabic Speaking Children in Kuwait," *J. Univers. Comput. Sci.*, vol. 19, no. 7, pp. 1003–1022, 2013.

[41] S. Dale, "Gamification: Making work fun, or making fun of work?," *Bus. Inf. Rev.*, vol. 31, no. 2, pp. 82–90, 2014.

[42] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work?–a literature review of empirical studies on gamification," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, 2014, pp. 3025–3034.

[43] S. Vandercruysse, M. Vandewaetere, F. Cornillie, and G. Clarebout, "Competition and students' perceptions in a game-based language learning environment," *Educ. Technol. Res. Dev.*, vol. 61, no. 6, pp. 927–950, 2013.

[44] G. G. Smith, M. Li, J. Drobisz, H. R. Park, D. Kim, and S. D. Smith, "Play games or study? Computer games in eBooks to learn English vocabulary," *Comput. Educ.*, vol. 69, pp. 274–286, 2013.

[45] H.-Y. Sung and G.-J. Hwang, "A Collaborative Game-based Learning Approach to Improving Students' Learning Performance in Science Courses," *Comput. Educ.*, vol. 63, pp. 43–51, 2013.

[46] A. Karkar, J. M. Al Ja'am, S. Foufou, and A. Sleptchenko, "An e-Learning Mobile System to Generate Illustrations for Arabic Text," *2016 IEEE Glob. Eng. Educ. Conf. EDUCON*, no. April, pp. 184–191, 2016.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

The Role of Cloud Computing in Learning from a Logistic and Strategic Vision

Case of Information Systems Management of a Private Moroccan University

S.Bourekkadi[1], S.Khoulji[1], O.Omari[2], K.Slimani[3], A.Gamal[1], M.L.Kerkeb[1]
[1]*Department of Networking and Information Systems, National School of Applied Sciences, UAE Tetouan, Morocco*
[2]*Department of English, Moulay Ismail University, Faculty of Arts and Humanities Meknes, Morocco*
[3]*Ibn Tofail University,Faculty of Sciences, Kenitra, Morocco*
acmlis.conference@gmail.com

ABSTRACT

This research focuses on the role of cloud computing from a logistic and strategic vision in organizations. Our research question focuses on understanding the difficulties and limitations that bring the support of steering direction function of information systems in the use of cloud computing in an organization. An exploratory study was conducted with an organization. The results are presented in this article in which we start with an introduction concerning cloud computing. This article includes also an understanding of this effective solution. Meanwhile, our work involves also citing changes of the case study after using cloud computing emphasizing on financial and material gains. In brief, this article discusses one of the most important subjects in the field of management, methodology and management. Especially through the presence of ERP within companies which becomes a compulsory element and not one of the choices.

Keywords-component; Cloud Computing, logistics, strategic, ERP, information systems, business applications, multimedia.

1 Introduction

The 21st century has experienced a new form that will rediscover the technology of information. This new concept is considered as a better source of solutions to the problems encountered in companies regardless of their size. Here, we are talking about “Cloud Computing”.

The principle of Cloud Computing is so simple, noting the following examples: storing photos, files, and data online (email or social networking website) instead of using a desktop. It is all about a cloud computing service. For instance, an organization that uses an online billing service instead of an in-house service, the online billing service is simply considered as a "cloud computing" service.The evolution of information technologies is significant in cloud computing. Cost reduction is a better argument for the providers who sell the cloud service. Users understand that corporate mobility benefits are the main advantage of a cloud computing movement. A study by Sand Hill found that 50% of respondents appreciate agility as the main reason for adopting cloud computing. A recent similar report according to

DOI: 10.14738/tmlai.54.3324
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3324

Information Week revealed that 65% of respondents also considered agility with respect to companies' needs as a motivation for cloud computing.

In fact, the main research question of this study is to know how can cloud computing help reduce costs while increasing the efficiency of the company? Hence, to have argumentative answers to this question we will analyze the research in three parts. At first, it seems wise to understand cloud computing. In a second step, we will use a survey on a well-known organization and analyze its existing state. This will allow us to discover in the third step, if the company decides to use cloud computing, what will be the benefits of this solution.

2 Defining Cloud Computing

In fact, Cloud Computing is the future of information technology. All major IT suppliers are investing their future revenues on cloud computing (Google, IBM, Microsoft, NASA, Adobe ...). Cloud Computing could lead to billions of dollars in energy savings.

Computing in a cloud or cloud computing is a major concept in the evolution of computing in the recent years. This concept refers to the use of computing and storage capacities of computers and servers distributed around the world and provided as a service through internet technologies to a multitude of external users.

The concept of Cloud Computing concerns the transfer of computer operations traditionally carried out on local computers to external servers. Thanks to cloud computing, users now have access to the various services that are managed online and become less complicated. The saved data is not on the local computer but in a cloud of remote servers interconnected by means of an excellent bandwidth essential to the fluidity of the system. That way, accessing the services is done via a web application that is standard and easily available.The tendency in information technology is "Cloud Computing" that characterizes computing in a network. In terms of outsourcing, companies use applications and data directly via the Internet through external interconnected servers of the organization. Then, companies find themselves free from a lot of information stored on their local posts.The cloud infrastructure contains several types of hardware components (servers, memory systems, network components, etc.), software application components (programs, services, protocols, etc.), and information.

A basic component of the cloud includes (Figure. 1):

- Technology architecture: It forms the first layer of the cloud platform (servers, operating systems, network tools, etc.).
- Application architecture: It is part of the functional layer and it includes all the business applications of the cloud platform.
- Information Architecture: This layer makes the cloud information available and ensures that it will be secure, consistent, and reliable.

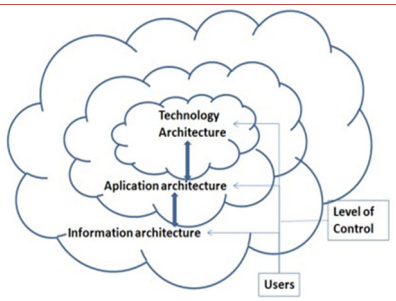


Figure 15 : The basic cloud computing architecture

According to the previous architecture scheme, a cloudplatform is able to provide:

- A completely independent infrastructure invisible to the final user;
- A platform capable of increasing its resources, if necessary, without the user being aware of it;
- An economical platform for the organization, reducing maintenance costs.

Meanwhile, the use of an ERP can be compatible with Cloud Computing if it is accessible in SaaS mode, and not in license mode. As a result, the company can access ERP applications directly on the Internet via remote servers where the data is stored.

Then, when we say ERP, we designate quite a module, among these modules we find:

1. Finance module.
2. Logistics module.
3. CRM module (GRC).

Indeed, the use of computing at the level of logistics becomes in our days among the principles of the IT. For this, we find a term that circulates between the computer scientists which is "*logistics*".

After all, how these two great jumps in technology can guide us to earn economically without touching the image of the organism, without losing the balance of the organization, and without becoming far from NICT especially that crises force companies to look for inexpensive solutions.

3 Does cloud computing propose savings?

While discussing the feasible savings that the Cloud offers, it is always necessary to do better with fewer budgets. Luckily, cloud computing services offer several financial benefits compared with other traditional systems.

3.1 Less meetings

The number of meetings and face-to-face meetings is significantly reduced through online service, collaborative writing, and discussion forums. For example, working on the same document remotely in collaboration with a team reduces the costs of moving people and saves time.

3.2 Knowledge management

The cloud is a good way to store data and share information within best practices, especially when a project is suspended or a department needs restructuring. Information here remains secure and centralized to the server. More importantly, it can be viewed and consulted by everyone whenever needed.

3.3 Coordination of shared services

The costs are reduced when the management committee is grouped together (logistics, marketing, finance ...)

3.4 Security access of the information

The cloud makes work easier either at home, on the road or anywhere. Productivity and efficiency are increased by using mobile devices.

3.5 Back on investment

The investment is justified when all people appreciate the service and facilitate their work which eventually makes the company benefit.

3.6 Security sharing of information

Users can easily share their confidential information without worrying in a secure and accessible environment from anywhere.

4 Exploratory study

A study was conducted in an organization in the North of Morocco within the IT systems department. The following information illustrates the current case study: (See Table 1)

4.1 Critical study

4.1.1 Materials:

Each year the organization spends an amount of 7,000,000 DHS of equipment for the renewal of their equipment. However, the problem is the difficulty of getting rid of the parts that are out of service which are also dangerous for the environment and take a lot of space in the organization. These storage workshops are useful for other features instead of hardware accumulation. The organization has 800 printers, 300 of them in the back office, most back office printers are not fully used and it is better to use instead network printers.

4.1.2 Quantitative of informatics park

Like most big organizations, the company suffers from the permanent increase of their computer center by adding new services; thus, it is necessary to have a solution before reaching the maximum limit of the size of computer center.

4.1.3 Software

The organization uses Windows operating systems, office, and antivirus as basic software. The price of the operating systems is included with the computer, which increases their prices. The company has 1000 fixed licenses of Microsoft Office but only 250 are used per day. Hence, the solution is to replace them with floating licenses (for example 400 licenses will be sufficient), this will of course reduce the cost of the purchased licenses.

Meanwhile, the software located in the stations requires a permanent update using a high bandwidth and a high rate of collisions (network link failure).

4.1.4 Consumable

The amount of annual energy consumption within the organization (electricity or fuel) is close to 800,000 DHS. The global crisis has forced companies to seek solutions to reduce consumable costs without increasing the quality of their service.

4.1.5 Maintenance

Maintenance is essential in the IT field but it is better to get rid of the daily problems in order to avoid reinventing the wheel.

Table 14 : The information about the experiment study

Study of materials	
Types of used materials	Desktop Pc, Laptop, Printer.
Annual purchase (amount)	7,000,000 DHS
Parts management (RAM, hard disk ...)	Stock in workshops: New part and Part out of service
Quantitative study of the computer park	
The state of computer center	The size increases permanently according to the portfolio of services
Software study	
Used software	Windows OS, Office, Antivirus
Difficulty of use	Permanent problem of operation
The amount of purchase	Included in the purchase of materials
Consumable study	
The consumption linked to the operation of the equipment	720,000 DHS
Diesel consumption due to technical team shifts	72,000 DHS
Maintenance study	
Contracts for maintenance	120,000 DHS
Inventory number	3600 per site
The most common problems	Material, software
The number of technicians at HELPDESK	17 on two sites

4.2 Results analysis

4.2.1 Minimizing the cost of material

By using Cloud Computing technology, we will no longer need a computer that costs 8,000 DHS; we will need only peripheral devices (case + screen) that cost just 1,700 DHS (4 times cheaper). This will reduce the cost of the annual purchasing by 78%. (See Fig.2)

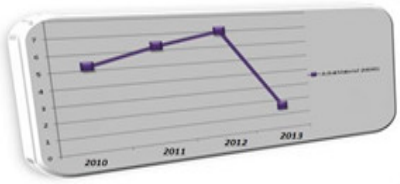


Figure 16. Evolution of the equipment cost

4.2.2 Consumable costs reduction

This is the most important thing when choosing a solution within the ISD, having an innovative project that gives added value to the company at a lower cost, at this moment there is no more project more efficient than Cloud Computing.

Cloud Computing offers a reduction in electrical consumables that can reach 50% using a Datacenter virtualization environment. Servers put on standby in charging State or Operating State. (see Fig. 3)

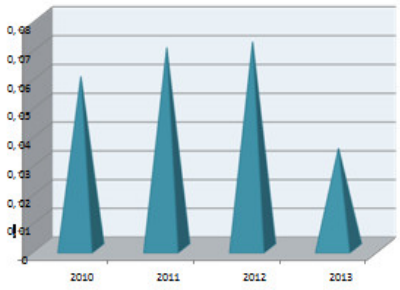


Figure 17. Consumable Equipment

4.2.3 Maintenance cost

Replacing computers with terminal devices (boxes) will eliminate maintenance contracts, the number of inventories per day will decrease, and software failure problems will disappear through the use of virtual desktops. Therefore, all this will increase the Quality of service offered by the company.

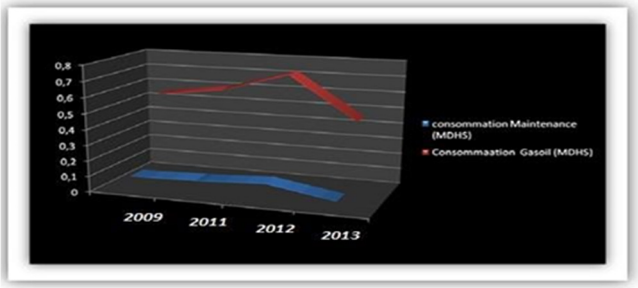


Figure 18 : Maintenance and fuel consumption

4.2.4 Elimination of parts' management and stock

By eliminating parts' management and the stock we will not need stock spaces that are useful in other functionalities; Moreover, we will save the dedicated amounts for the recycling of non-service parts.

Minimizing human and technical resources by using Cloud Computing, we will not need a large number of technicians because it will automatically decrease in relation to reduced technical resources. Virtualization will serve the solution of printing problems:

Printing is a permanent and indispensable service in the organization; precisely the printing of customer bills that are available and each incident gives a bad image within the organization.In implementing the Cloud Computing solution, we noticed that within the back office, the number of printers is 300 with only 50 printers used per day, plus the cost of paper and ink purchased in a random way.

For this reason, it has been proposed to set up a printer server within the Back office; a print server is a server that allows one or more printers to be shared between multiple users (or computers) on the same computer network.

This brings considerable and significant savings such as:

- Space saving

It is no longer necessary to buy one printer per user. This saves space, electricity, and maintenance.

- Quick setup

Add a printer in few minutes. No need to cut your network connection.

- Flexibility, no queue or delay

Networking one or more printers can provide you with increased flexibility.

- Increased Cables Distance

The maximum recommended distance for a USB or Parallel printer is 5 meters. With the print server, you extend the connection length to 100 meters by connecting a printer directly, or 200 meters through a Switch or a network hub.

- Connection by cable category 5

Some modern buildings already have this type of wiring installed originally. It is easy to install it in case it is not installed.

- Model one or more ports according to your needs

If you need to connect more than one printer to your network, you can use multiport patterns. In this case, only one network port will be used.

5 Results

From the second year, students will be required to do practical work, download and buy Photoshop software, animation, flash, image games, and editing videos. Therefore, to use the 3D software it is better to have PCs with high quality and high capacity, such as i7 for example which can work with this software. So, with the solution Cloud Computing, it will be possible to buy a 700 DH case with screens of 400 DH that worth about 1,100 DH for each with the Cloud installed as a service. Hence, this will be much better than buying a PC of 17,000 DH.

For example, buying 10 licenses for 10 PCs will cost much more than installing the Cloud and purchasing only 5 licenses to work with the same number of PCs. Moreover, it will be possible to install, work, and use this software as a service and not as a product. We can also pay companies that will provide us with this monthly or annual product.

In addition, students who have not yet finished their work at school will be able to continue their work at home with the same application. Therefore, the access to the software will be online, that means all the work will be in cloud. This is why the ideal solution for like these issues is Cloud Computing.

5.1 The financial gain

According to the latest CSC survey, 82% of companies that have implemented a Cloud project have made financial gains. Thanks to the benefits provided to the company by the Cloud Computing solution.

A survey was done by the CSC based on a collection of 3465 users using Cloud Computing solution features due to the following reasons:

Instead of buying expensive hardware or software to use a particular program, Cloud Computing gives access to programs or applications centralized in servers and accessible to all users. The cost savings of not having to pay for equipment and programs that will quickly become obsolete will be extremely helpful to your budget.

Some Cloud Computing services give the possibility to increase the processing power. Even when there is a need for increased processing capacities for a short period of time, it is possible to do it without having to invest heavily in expensive equipment based on Data center's virtualization technologies.

Large companies can use the cloud to endow the entire organization with tools that previously had to be installed on each computer, which wasted IT some time and was expensive. Meanwhile, companies will not need to renew their hardware based on recommended operating software and applications that save annual hardware purchases, inventory management, and recycling expenses for non-service materials.

5.2 The material gain

Cloud computing is a concept of deportation on remote servers of computer treatments traditionally performed on local computers, which does not recommend new software. With the Cloud Computing the client receives only an interface of the application that he/she uses and that is executed in the server. Cloud computing does not recommend high-performance hardware to customer support due to the installed software (an operating system + a web browser). The purchase of hardware is an indispensable task in big companies that apply the international standards of quality of services. This consists of the renewal of clients, servers etc., but it is not the same with Cloud Computing. The hardware purchase does not take into account customer support; there is a need of renewal of the servers in term of their life cycle and performance. However, despite the advantages, it is necessary to change the computer equipment every 4 years according to the age of depreciation.

6 Conclusion

To conclude, it was also possible to realize that the “IS” is an area in which it is necessary to work every day if we want to be able to keep up to date and to deepen one's knowledge. Indeed, new versions of the MSI tools come out regularly, each time with new features or improvements compared to the previous version. In addition, IT is a sector that is constantly evolving. Consequently, if we do not want to be overtaken too quickly, we have to keep up to date and for that there is only one solution, it is to work in a self-taught way very often. Cloud Computing² is nowadays the most real cloud ever. It is a futuristic technology but it is current! If we were to summarize in a sentence what Cloud Computing is, we could say that it is a concept of deporting resources on remote servers in contrast to traditional hosting on the user station.

ACKNOWLEDGMENT

This work was supported by the grant of National Center for Scientific and Technical Research (CNRST-Morocco): No. 6UAE2015 and 757UIT for the first and the fourth authors.

We would like to express our gratitude and appreciation to Mrs. Nadia BOUMAZZOUGH, M. Marouan ELMHOUJEB, and M. Abdeljabar MAHDAR for their great inspiration, moral supports and their guidance.

REFERENCES

[1] A. Sirotkin, Kaisa Koskela-Huotari, K. Karppinen, J. Del Ser, B. McCabe, "Differentiation in the Cloud: Methodology for Integrating Customer Values in Experience Design", Software Engineering and Advanced Applications (SEAA) 2012 38th EUROMICRO Conference on, pp. 453-459, 2012.

[2] Andreas Papadopoulos, Dimitrios Katsaros, "A-Tree: Distributed Indexing of Multidimensional Data for Cloud Computing Environments", Cloud Computing Technology and Science (CloudCom) 2011 IEEE Third International Conference on, pp. 407-414, 2011.

[3] Ximeng Liu, Robert H. Deng, Kim-Kwang Raymond Choo, Jian Weng, "An Efficient Privacy-Preserving Outsourced Calculation Toolkit With Multiple Keys", Information Forensics and Security IEEE Transactions on, vol. 11, pp. 2401-2414, 2016, ISSN 1556-6013.

[4] Jih-Sheng Chang, Ruay-Shiung Chang, "A performance estimation model for high-performance computing on clouds", Cloud Computing Technology and Science (CloudCom) 2012 IEEE 4th International Conference on, pp. 275-280, 2012.

[5] Chuang Deng, Junyong Liu, Yang Liu, Zhen Yu, "Cloud computing based high-performance platform in enabling scalable services in power system", Natural Computation Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) 2016 12th International Conference on, pp. 2200-2203, 2016.

[6] Dinh, H. T., Lee, C., Niyato, D., & Wang, P. “A survey of mobile cloud computing: architecture, applications, and approaches”. Wireless communications and mobile computing, 13(18), 1587-1611, 2013.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Detection of Damage in Rail Head by Using SAFE Method

Sara.Teidj, Abdellah. Driouach, Abdellatif. Khamlichi
Systems of Communications and Detection Laboratory
Faculty of Sciences, Tetuan, Morocco
sara.teidj@gmail.com, adrioauch@hotmail.com, khamlichi7@yahoo.es

ABSTRACT

The first cause of train derailment incidents is due to the transverse defect which occurs in the railhead. It is typically an open or internal crack developing in a plane which is generally perpendicular to the direction of the rail. The most method used today of rail inspection is based on ultrasound .The Ultrasounds based testing is performed according to the excitation-echo procedure. It is conducted conventionally by using a contact excitation probe that rolls on the rail head or by a contact-less system using laser as excitation and air coupled acoustic sensors for wave reception [1].

In this work, Propagation of guided elastic waves in a rail has been modelled by using the semi-analytical finite element method [2]. We have demonstrated the existence of several frequency windows that can be used to excite propagating modes with high deformation in the railhead.

Keywords-rail; waveguide; pulsed laser; air-coupled; safe method; laser vibrometry.

1 Introduction

With increasing railways traffic and augmenting speed of modern trains, maintenance of rails constitutes a major concern of safety for this form of transportation. The objective is to impede catastrophic failures that can threaten human lives and provoke huge economic loss [3].

Rail is the major cause of train accidents [4]. Almost 70% of the causes of train derailments are attributed to the transverse defect appearing in the railhead [5]. Current inspection techniques are mainly based on electromagnetic induction [6] or ultrasound [7].

Since an analytical method cannot be used to modeling wave propagation in waveguides of any arbitrary cross-sections a new numerical method has been developed. It is the Semi-Analytical Finite Element (SAFE) method [8]. This method has been successfully used for several structures and in particular in the case of any cross-section beams such as rails. The method requires the finite element discretization of the cross- section of the waveguide, and assumes harmonic motion along the wave propagation direction.

We present the SAFE method and we will apply it through the Comsol Mutiphysics software to model the elastic waveguides in a rail. We illustrate the study of propagative modes that can account for the presence of transverse defects in the railhead.

DOI: 10.14738/tmlai.54.3325
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3325

2 Elastic Waveguides in a Rail

2.1 Propagation of elastic waveguides

Let us consider an elastic waveguide with waves propagating along the direction x_3 with the wave number k and frequency ω. The cross-section of the waveguide is in the plane (O,x_1,x_2).

The equation of motion is written:

$$C_{iplm}\frac{\partial^2 u_l}{\partial x_p \partial x_m} = \rho \frac{\partial^2 u_i}{\partial t^2} \qquad i,p,l,m = 1,2,3 \tag{1}$$

Where C_{iplm} are the terms of the elasticity tensor, ρ the density of the material, u_i the components of the displacement vector and t the time.

In harmonic mode the general solution of the waveguide is in the form

$$u_i(x_1,x_2,x_3,t) = U_i(x_1,x_2)\exp\left[j\left(kx_3 - \omega t\right)\right] \quad i = 1,2,3 \tag{2}$$

Where k the wave number, ω pulsation and U_i amplitude.

By substituting equation (2) in equation (1), it comes

$$C_{iplm}\frac{\partial^2 u_l}{\partial x_p \partial x_m} + \rho\omega^2 u_i = 0 \qquad i,p,l,m \in \{1,2,3\} \tag{3}$$

By explaining the calculation of derivation with respect to x_3 we obtain:

$$C_{iplm}\frac{\partial^2 U_l}{\partial x_p \partial x_m} + jk\left(C_{i3lp} + C_{ipl3}\right)\frac{\partial U_l}{\partial x_p} - k^2 C_{i3l3}U_l + \rho\omega^2\delta_{il}U_l = 0 \tag{4}$$

with $\quad i,l \in \{1,2,3\} \quad p,m \in \{1,2\}$

To equations (4), we must add the boundary conditions on the free boundary of the right section which are written:

$$T_i = \sigma_{ip}n_p = C_{iplm}\frac{\partial U_l}{\partial x_m}n_p + jkC_{ipl3}U_l n_p = 0 \tag{5}$$

The phase velocity is defined by:

$$c_p = \frac{\omega}{k} \tag{6}$$

Where one takes the real wave numbers knowing that in general they are complex.

If the wave number is pure imaginary, the wave is said to be evanescent and does not propagate.

When $real(k) \neq 0$,the propagation is possible, then the imaginary part $imag(k)$ introduces attenuation and the wave is damped.

2.2 Calculate the Dispersion curves in a rail by Comsol Multiphysics software

We present a method which makes it possible to use Comsol Multiphysics software to perform the calculation of dispersion curves. This method uses the PDE mode (Partial Differential Equation), which allows to the user to define the problem by writing the differential equations to the partial differentials that govern it.

The PDE coefficient form interface of Comsol Multiphysics makes it possible to solve a boundary problem governed by the system of differential equations with partial derivatives having the following [9]:

$$\nabla.\left(c\nabla\vec{u}+\alpha\vec{u}\right)-\beta\nabla\vec{u}-a\vec{u}+\lambda d_a\vec{u}=0 \qquad dans \quad \Omega \tag{7}$$

$$\vec{n}.\left(c\nabla\vec{u}+\alpha\vec{u}\right)=0 \qquad sur \quad \partial\Omega \tag{8}$$

where $\vec{n}$ the unit normal vector external to the domain and $\vec{u}$ the unknown field.

A reformulation of the SAFE problem defined by the system of equations (4) has been proposed by Predoi [10]. Equation (7) can be rewritten in the following explicit form:

$$C_{iplm}\frac{\partial^2 u_p}{\partial x_l \partial x_m}+(\alpha_{ipl}-\beta_{ipl})\frac{\partial u_p}{\partial x_l}-a_p u_p+\lambda d_{ip}u_p=0 \tag{9}$$

Similarly the Neumann condition defined by equation (8) is identified by

$$C_{iplm}\frac{\partial u_p}{\partial x_m}n_l+\alpha_{ipl}u_p n_l=0 \tag{10}$$

In order to rewrite the system of equations (9) and (10) the set of the following matrices is introduced:

$$d_a=\begin{bmatrix} O & D \\ M & O \end{bmatrix}; c=\begin{bmatrix} C & O \\ O & O \end{bmatrix};\ \alpha=\begin{bmatrix} O & jA \\ O & O \end{bmatrix} \tag{11}$$

$$\beta=\begin{bmatrix} O & -jB \\ O & O \end{bmatrix}; a=\begin{bmatrix} M & O \\ O & M \end{bmatrix}$$

The submatrices are given by

$$A=\begin{bmatrix} C_{15} & C_{14} & C_{13} \\ C_{65} & C_{64} & C_{63} \\ C_{65} & C_{64} & C_{63} \\ C_{25} & C_{24} & C_{23} \\ C_{55} & C_{54} & C_{53} \\ C_{45} & C_{44} & C_{43} \end{bmatrix}; B=\begin{bmatrix} C_{51} & C_{56} & C_{55} \\ C_{56} & C_{52} & C_{54} \\ C_{41} & C_{46} & C_{45} \\ C_{46} & C_{42} & C_{44} \\ C_{31} & C_{36} & C_{35} \\ C_{36} & C_{32} & C_{34} \end{bmatrix}$$

$$C=\begin{bmatrix} C_{11} & C_{16} & C_{16} & C_{12} & C_{15} & C_{14} \\ C_{61} & C_{66} & C_{66} & C_{62} & C_{65} & C_{64} \\ C_{61} & C_{66} & C_{66} & C_{62} & C_{65} & C_{64} \\ C_{21} & C_{26} & C_{26} & C_{22} & C_{25} & C_{24} \\ C_{51} & C_{56} & C_{56} & C_{52} & C_{55} & C_{54} \\ C_{41} & C_{46} & C_{46} & C_{42} & C_{45} & C_{44} \end{bmatrix} \tag{12}$$

$$D=\begin{bmatrix}-C_{55} & -C_{54} & -C_{53}\\ -C_{45} & -C_{44} & -C_{43}\\ -C_{35} & -C_{34} & -C_{33}\end{bmatrix};$$

$$M=\begin{bmatrix}-\rho\omega^2 & 0 & 0\\ 0 & -\rho\omega^2 & 0\\ 0 & 0 & -\rho\omega^2\end{bmatrix};O=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

Where the C_{pm} $p,m=1,...,6$ are the contracted coefficients of the elasticity tensor and $\lambda=k$.

In the case of a linear and isotropic elastic material of Young's modulus E and Poisson's ratio ν, the matrices A, B, C and D are written

$$A=\frac{E}{(1+\nu)(1-2\nu)}\begin{bmatrix}0 & 0 & \nu\\ 0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1-\nu\\ 1-2\nu & 0 & 0\\ 0 & 1-2\nu & 0\end{bmatrix} \tag{12}$$

$$B=\frac{E}{(1+\nu)(1-2\nu)}\begin{bmatrix}0 & 0 & 1-2\nu\\ 0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1-2\nu\\ \nu & 0 & 0\\ 0 & \nu & 0\end{bmatrix}$$

$$C=\frac{E}{(1+\nu)(1-2\nu)}\begin{bmatrix}1-\nu & 0 & 0 & \nu & 0 & 0\\ 0 & 1-2\nu & 1-2\nu & 0 & 0 & 0\\ 0 & 1-2\nu & 1-2\nu & 0 & 0 & 0\\ \nu & 0 & 0 & 1-\nu & 0 & 0\\ 0 & 0 & 0 & 0 & 1-2\nu & 0\\ 0 & 0 & 0 & 0 & 0 & 1-2\nu\end{bmatrix}$$

$$D=\frac{E}{(1+\nu)(1-2\nu)}\begin{bmatrix}2\nu-1 & 0 & 0\\ 0 & 2\nu-1 & 0\\ 0 & 0 & 2\nu-1\end{bmatrix}$$

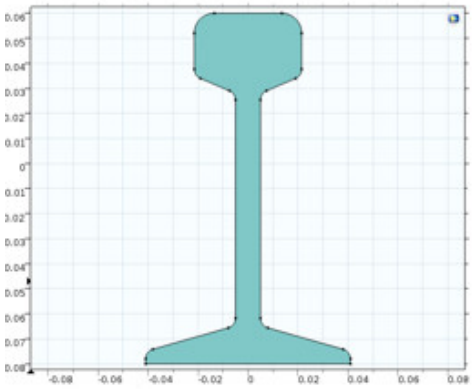
4 Results and discussion

We apply the SAFE approach to determining the dispersion curves in the case of a rail. The knowledge of dispersion curves presents a great interest in the domain of the CND which concerns frequencies up to $50\,kHz$ [11].

We consider a homogeneous and isotropic rail whose material properties are: $\rho=7800\,kg/m^3$; $E=2\times10^{11}\,Pa$ and $\nu=0.3$

We consider a transverse defect which is modeled by a loss of section in the form of a rectangle whose dimensions are: $0.01m\times0.02m$.

Figure 1 and 2 show the section of the rail without defect and the rail with defect.



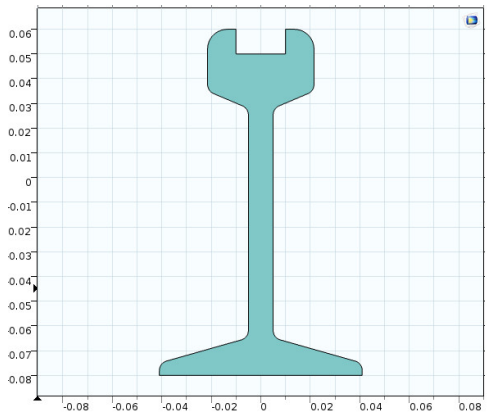


Figure 2: European Rail Geometry Vignole; with defect in railhead

Figure 3 and 4 show the mesh used in the SAFE method for the rail without defect and the rail with defect.

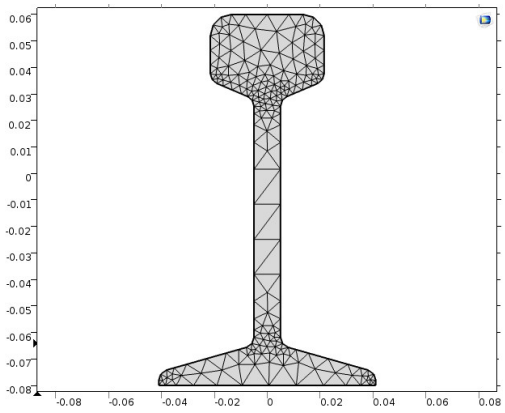


Figure 3: Meshing of European rail Vignole UIC 18.3

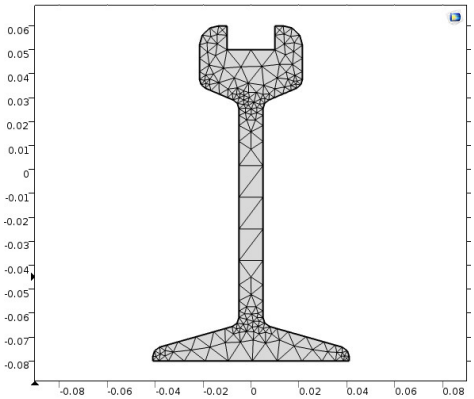


Figure 4: Meshing of European rail Vignole UIC 18.3with defect in railhead

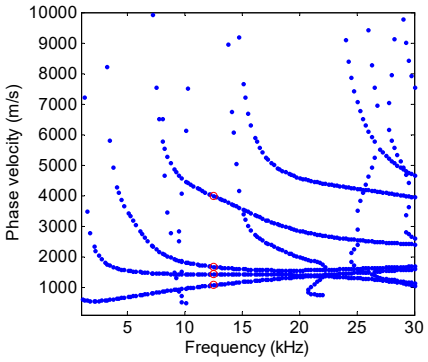


Figure 5: Dispersion curves in terms of phase velocity for the intact rail; the points red are for $12.5kHz$

The dispersion curves in terms of phase velocity are shown in figure 5 in the case of the intact rail for frequencies up to the value $30\,kHz$.

This figure shows a frequency window around the frequency $12.5kHz$ that can be used to excite the rail. The analysis of the complex wave numbers showed that the attenuation increases strongly with frequency.

The knowledge of high-frequency attenuation is essential in order to identify the low-loss propagation modes that can be effectively used to operate detection.

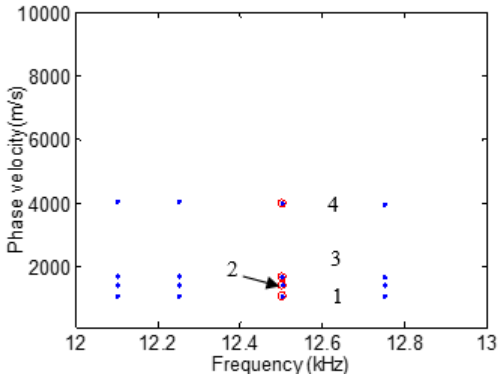


Figure 6: Zoom of dispersion curves in terms of phase velocity in $[12, 13]\text{kHz}$

Figure 6 shows a zoom on the dispersion curves in the frequency interval .The first four modes of the cross section are numbered from the low to the up.

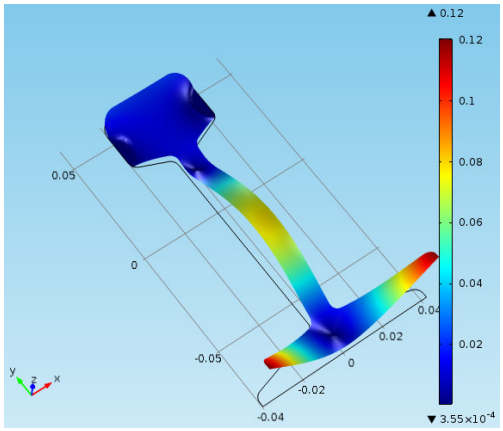


Figure 7: Intact rail; Displacement standard for $f = 12.5kHz$ **and** $k = 54.994\,/\,m$

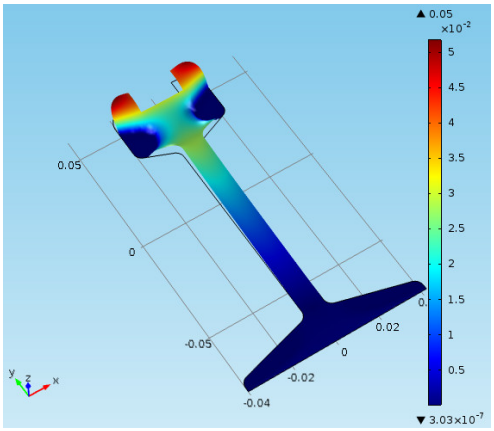


Figure 8: rail with defect; Displacement standard for $f = 12.5kHz$ **and** $k = 85.506\,/\,m$

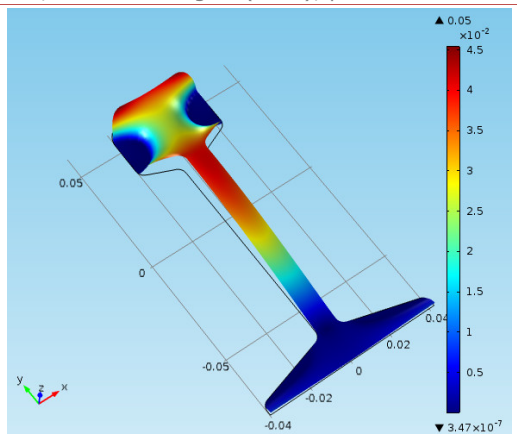


Figure 9: Intact rail; Displacement standard for $f = 12.5kHz$ and $k = 72.347/m$

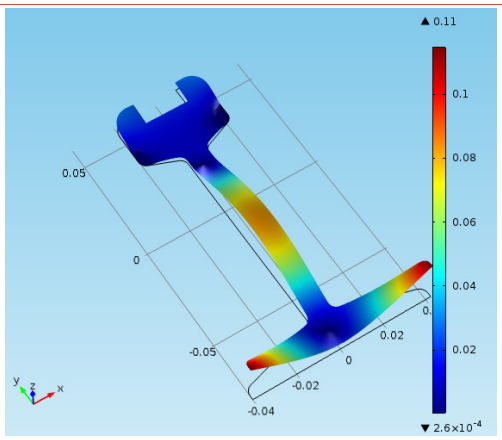


Figure 10: Rail with defect; Displacement standard for $f = 12.5kHz$ and $k = 55.513/m$

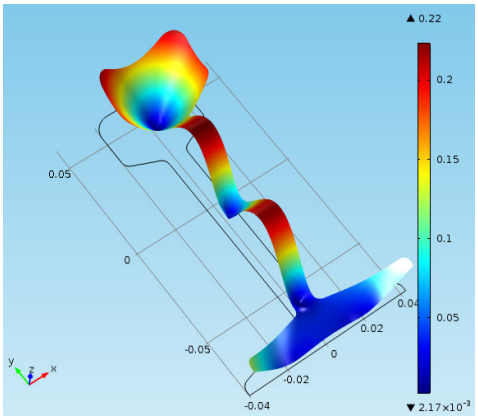


Figure 11: Intact rail; Displacement standard for $f = 12.5kHz$ and $k = 19.645/m$

$k = 19.645/m$

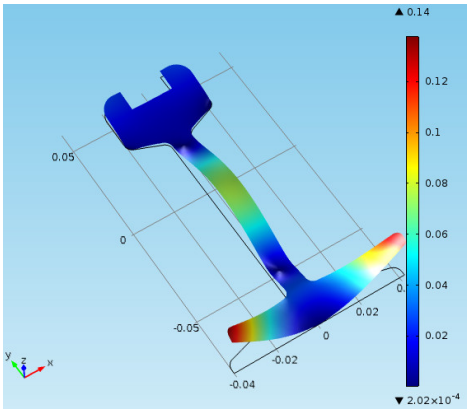


Figure 12: Rail with defect; Displacement standard for $f = 12.5kHz$ and $k = 47.025/m$

$k = 19.645/m$

This information can be exploited in practice in CND tests to target defects according to their position in the rail section.

5 Conclusion

The knowledge of these modes of propagation is very important in order to be able to design the ultrasonic wave excitation and detection system. This is to find the most suitable configuration for the type of defect.

REFERENCES

[1] B.D. Jeffery, M.L. Peterson, "Assessment of Rail Flaw Inspection Data," A Center of Excellence for Rural and Intermodal Transportation, 1999.

[2] I. Bartoli, “Structural health monitoring by ultrasonic guided waves,” PhD Thesis, University of California, San Diego, 2007.

[3] U. Zerbst, R. Lundén, K.O. Edel and R.A, Smith, “Introduction to the damage tolerance behaviour of railway rails,” Engineering Fracture Mechanics,vol. 76, pp. 2563-2601, 2009.

[4] D.F. Cannon, K.O. Edel, S.L. Grassie, K. Sawley, “Rail defects: an overview,” Fatigue and Fracture of Engineering Materials and Structures, pp. 865-887, 2003.

[5] K. Sawley, R. Reiff, “Rail Failure Assessment for the Office of the Rail Regulator,” Report No P-00-070 produced by Transportation Technology Center Inc, Pueblo, Colorado, USA, 2000.

[6] C. Duverger, B. Prazil , “Un nouvel engin d’auscultation des rails à la SNCF,” Revue Générale des Chemins de Fer. Paris: Bordas-Dunod,pp. 1-12, 1984.

[7] Y. Fan, S. Dixon, R.S.Edwards, X. Jian, “Ultrasonic surface wave propagation and interaction with surface defects on rail track head,” NDT & E International,pp. 71-7,2007.

[8] T. Hayashi, W. J. Song, J. L Rose, “Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example,” Ultrasonics,pp. 175-183,2003.

[9] COMSOL User's Guide and Documentation version 4.3a. COMSOL AB.

[10] M.V. Predoi, M. Castaings, B. Hosten, C. Bacon , “Wave propagation along transversely periodic structures,” Journal of the Acoustical Society of America,pp. 1935-1944,2007.

[11] F.L. di Scalea., J. McNamara, “Ultrasonic NDE of railroad tracks: air-coupled cross-sectional inspection and long-range inspection,” Insight - NDT& Condition Monitoring, vol. 45,pp. 394-401, 2003.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Building a Smart Interactive Kiosk for Tourist Assistance

Hanane Amessafi[1], Reda Jourani[2], Adil Echchelh[1], and Houssain Oulad Yakhlef[3]
[1]*Laboratory of electrical engineering and energetic systems*
Faculty of Sciences,Kenitra, Ibn Tofail University, Kenitra, Morocco
[2]*Information Systems Engineering Group*
Polydisciplinary Faculty of Tétouan, AbdelMalek Essaâdi University, Tétouan, Morocco
[3]*Modelling and information theory Group*
Polydisciplinary Faculty of Tétouan, AbdelMalek Essaâdi University, Tétouan, Morocco
hanane.amessafi@gmail.com, reda.jourani@yahoo.fr, adilechel@gmail.com, houlad10@gmail.com

ABSTRACT

The tourism sector in morocco is increasing rapidly, is well developed, with a strong tourist industry focused on the country's coast, culture, and history. Morocco is the most politically stable countrie in North Africa, which has allowed tourism to develop. There are some tourist attractions spread over a wide geographical area, which are only visited by a few people at specific times of the year. Additionally, having human tourist guides everywhere and speaking different languages is unfeasible. This work deals with the build of a smart interactive kiosk for tourist assistance. In order to do this, it is necessary to start with an Automatic Language Recognition to recognize the language used by the tourist, then extracting keywords from what the tourist said by using Automatic Speech Recognition techniques among others, and finally the system analyses and ansers to the tourist query by suggesting a set of related informations.

Keywords- Automatic Language Recognition, Automatic Speech Recognition, keyword extraction, question answering, Gaussian Mixture models, tourism

1 Introduction

In order to offer an attractive service for tourists visiting the kingdom of Morocco, we aim to build smart interactive kiosks for tourist assistance. These kiosks try to help tourists to find points of interest. Tourism is one of the most effective money-moving mechanisms, redistributing wealth, and promoting local economies. It brings income into a community that would otherwise not be earned[1]

Information systems are always dynamic, so learning adapts to new models of functioning by focusing first on automatic speech recognition while exploiting the news and the latest kinds of technologies, by analyzing the keywords from what the tourist said.

A recognition system includes several components:

A parameter extraction module, a matching block, a module for normalizing the match scores and a decision module.

DOI: 10.14738/tmlai.54.3326
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3326

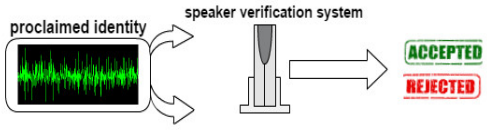


Fig 19 of a speaker verification system

Innovation consists in developing a multidisciplinary approach, bordering on the sciences of engineering and the sciences of analysis and treatment, favoring the interactions of the tourist with his need, based on field study and modelling. We will describe social-technical systems, in other words, the interdisciplinary methods of different problems, text comprehension, analysis of the blogosphere or social networks as well as artificial life and robotics: artificial learning that can be Endowing autonomous systems, which are inspired by natural behaviors without forgetting the integrative neurosciences: brain computing, perceptual-motor modelling ... which focus on the reactions of the subjects and the learning within the modelling of the action.

Professionals in this field confirm that 90% of the knowledge are stored in the heads, while the 10% are recorded in books and computers, the latter can search, store and process knowledge, but the importance is in the level of knowledge transfer. Much information and knowledge is on the web, but the manipulation of this knowledge is not well developed. Artificial Intelligence has its role to play.

The expert system avoids having to write new programs to reinject the information: thanks to its acquisition module, new data can be incorporated during use. It is designed to help a user in a particular field to find the appropriate solution for his inquiry, and this, of course, in the current state of specialised knowledge.

In this communication, we will describe two fundamental points: first, the questioning of "behaviorism", and secondly, the possibility of constructing models integrating data from neurosciences and cognitive psychology by using simulations and modeling Of information technology.

The paper is organized as follows. After an overview on Automatic Language Recognition in section 2, we describe our Methodology in section 3. The experimental results and their discussions are then presented in section 4

2 Maintaining the Expert System and Automatic Language Recognition

The elaboration of an expert system is based on an approach essentially borrowed from the humanities. Surely, there is no ready-made method to define the different strategies of the experts. But, generally, the designer seeks to draw three levels within the mass of knowledge. See the table below.

Table 1 Levels of Knowledge

Structuring level	*Conceptual level*	*Cognitive level*
Touches algorithmic procedures in the field of multilingual systems for automatic recognition of the language	Represents knowledge in order to represent the concepts usually used by the specialist.	Gathers a large amount of raw knowledge related to the field of touristic communication.

For a system of rules, it is estimated that by the help of interviews, we can generate a very limited number of rules per day.

The causes are as follows:

The engineer is not a specialist in the field of expertise, but must learn a minimum knowledge base to communicate with the expert.

The expert does not generally think in terms of broad principles but in terms of typical situations and classical events.

It is already difficult to develop a framework (notations) to express knowledge, even on paper.

To exceed the limits of the expert system it is necessary to:

Better model the reasoning.

Better model the field.

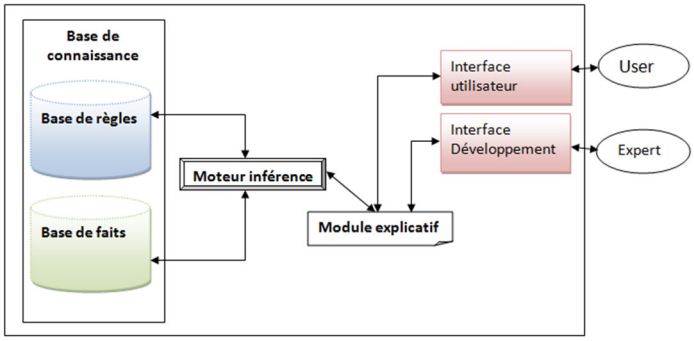


Figure 20 Design of an expert system

3 Methodology

Knowledge engineering would evoke techniques for manipulating knowledge on computer.

The effort is focused on:

Identification & Acquisition :

Techniques & Tools (documents, interviews ...)

Modeling & Formalization: :

Language of modeling (field, task...)

Methodologies (KADS, KOD...)

Language of formal representation

Use

Surfing, inference, explanation...

Maintenance and management

Intervention of several disciplines

Artificial intelligence

Information systems & documentaries

Software engineering

Object Oriented Programming, Agent

Logic Programming

Human machine interfaces

Linguistics (Natural language, Terminology)

Logic (Formal logical reasoning)

Psychology

Ergonomics (HMI, user approach)

Pedagogy & Teaching

Philosophy (with ontology)

Semiotics (Signs and Representations)

Sociology & Anthropology (collective)

4 Case Study and Results

A statistical and scientific study has been carried out in order to promote our proposed solution in the field of tourism, in order to have and ensure a high level of communication between the tourist and the various receptionists. Here is what presents as results:

Taking into consideration the language of tourists can give more to the national economy.

Extracting keywords from what the tourist said.

Proposing responses to the tourist in relation to the extracted keywords.

So it is possible to say and to confirm that adaptive communication to the language of a tourist has an impact on the quality of service provided, the same method used by professionals "texts translated in several languages" does not meet their need, from wich comes the need for innovation of an INTELLIGENT MULTILINGUAL TOURIST ASSISTANCE SYSTEM.

Besides, the absence of ‘face-to-face’ communication remains an obstacle to the success of the Vision 2020 objectives. Therefore, the establishment of a communication and tourist assistance system will enable:

Meeting the needs of foreign tourists.

The establishment of a closer working environment to the reality.

5 Conclusion

The idea is to set up a computer terminal for tourists in order to provide them with certain information. Starting with an "automatic recognition of the language" that was used by the tourist, then extract keywords from what the tourist says ("automatic speech recognition"), and then propose responses to the tourist In relation to the extracted keywords. so our system is able to correctly inform tourists about interesting tourism topics and to propose other related information about any touristic issue.

For example, information will be given after saying the name of a city or a monument. Accommodation or catering sites will be indicated if the tourist says "hotel", "guest house", or "restaurant "..., we can treat the means of transport.

Artificially intelligent agents, and other technological breakthroughs will help tourists to complete more transactions in less time and gather more relevant information about their destinations in real-time [1]

REFERENCES

[1] Garrido, P., Barrachina, J., Martinez, F., Seron, F., 2017. Smart tourist information points by combining agents, semantics and AI techniques. Comput. Sci. Inf. Syst. 14, 1–23. doi:10.2298/CSIS150410029G

[2] Towards a cognitive approach to human machine cooperation in dynamic situations : JEAN-MICHEL HOC

[3] DePaulo, J. Lindsay, B. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. 2003. Cues to deception. Psychological Bulletin, 129(1):74—118.

[4] Computer Simulation of Human Thinking : Allen Newell, Herbert A.Simon

[5] The Prospects for Psychological Science in Human-Computer Interaction : Allen Newell, Stuart K. Card

[6] Youcef B.C., Elemine Y.M., Islam B., Farid B. (2017) Speech Recognition System Based on OLLO French Corpus by Using MFCCs. In: Chadli M., Bououden S., Zelinka I. (eds) Recent Advances in Electrical Engineering and Control Applications. Lecture Notes in Electrical Engineering, vol 411. Springer, Cham

[7] Patel, K., Prasad, R.K.: Speech recognition and verification using MFCC & VQ. Int. J. Emerg. Sci. Eng. (IJESE) 1(7) (2013). ISSN: 2319–6378

[8] Jourani, R. and Daoudi, K. and Andre-Obrecht, R. and ´Aboutajdine, D., "Large Margin Gaussian mixture models for speaker identification," in Proc. of Interspeech, 2010, pp. 1441–1444

[9] Achraf Ben Romdhane, Salma Jamoussi, Abdelmajid Ben Hamadou, Kamel Smaili Phrase-Based Language Model in Statistical Machine Translation ;International Journal of Computational Linguistics and Applications, Alexander Gelbukh, 2016 Journal articles

[10] Jourani, R., Daoudi, K., Andr´e-Obrecht, R., and Aboutajdine, D. (Online First). Discriminative speaker recognition using Large Margin GMM. Journal of Neural Computing &Applications, doi :10.1007/s00521-012-1079-y.

[11] Jourani, R., Daoudi, K., Andr´e-Obrecht, R., and Aboutajdine, D. (2011). Speaker verification using Large Margin GMM discriminative training. In Proc. of ICMCS, page 1–5

[12] M.E. Al-Ahdal and N.M. Tahir, "Review in sign language recognition systems," in ISCI'12 : IEEE Symposium

[13] A.K. Sahoo, G. S. Mishra, and K. K. Ravulakollu,"Sign language recognition : State of the art," ARPN Journal of Engineering and Applied Sciences, vol. 9,no. 2, pp. 116–134, 2014

[14] M. Koppel, S. Argamon, and A. Shimoni. 2002. Automatically categorizing written texts by author gender. Literary and Linguistic Computing, 4(17):401– 412.

[15] Karn ("Design And Evaluation Of A Phonological Phrase Parser For Spanish Text-To-Speech", Fourth International Conference on Spoken Language, Oct. 1996).*

[16] T.M. Derwing, M.J. Munro, and M. Carbonaro, “Does popular speech recognition software work with ESL speech?”, TESOL Quarterly 34, 592-603, 2000. [2] D. Coniam, “Voice recognition software accuracy

[17] T. Chen, C. Huang, E. Chang, and J. Wang, “Automatic accent identification using gaussian mixture models,” in Automatic Speech Recognition and Understanding, 2001. ASRU’01. IEEE Workshop on. IEEE, 2001, pp. 343–346.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Generic Platform to Generate Educational Components

[1]Yassine Aarab, Noura Aknin, [2]Abdelhamid Benkaddour
[1]Information Technology and Modeling Systems Research Unit
Faculty of Science,Abdelmalek Essadi University
Tetuan, Morocco
[2]Computer Science, Operational Research and Applied Statistics Laboratory
Faculty of Science, Abdelmalek Essadi University
Tetuan, Morocco
yacinorock@gmail.com, aknin@ieee.com; ham.benkaddour@yahoo.fr

ABSTRACT

In this research, we will present all the concepts necessary for the implementation of a warehouse of educational components. We define three classes of educational components: Learning components, services components and technical components. We built a metadata schema to describe them. We will study the warehouse architecture and how the actors interact with it. We will also see the features implemented at the indexing and search system that integrates. The aim is to clarify the strategic solutions we have adopted to create a generic platform, adaptable and scalable. The instantiation of this platform provides the Educational Component Repository.

Keywords: E-Learning, Component Software, Metadata, Warehouse.

1 Introduction

A training curriculum well built is realized by the composition of educational entities of granularity varied. In the literature, these entities are designated, sometimes with confusion, by the terms components, resources, objects or materials. Based on the definitions the more widespread, we want to propose the rules of distinction between these different terms.

Bourda uses the term educational object defined as "any entity, Digital or not, which can be used, re-used or referenced in any activity related to teaching or the learning" [1][2]such as transparencies, course notes, web pages, of simulation software, etc.

The definition more recovery by the Community has been given by the Sub-group P1484.12, belonging to the group of work (LTSC Learning Technology Standards Committee) of the IEEE responsible for describing the educational objects. This group defines a educational object, as "an entity on IT support or not, which can be used, re-used or referenced during the activities of computer-assisted learning (education - intelligent- assisted by computer, learning environments, interactive distance education system, collaborative learning environments.)" [3].

It postulates that everything that can be used in the teaching by computer is considered as an educational object. A course digital, a QCM, a book are educational objects if they are used or referenced in a computerized course. By contrast, transparencies or exercises scanned, used in a traditional course, are

DOI: 10.14738/tmlai.54.3327
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3327

not considered as educational objects because they are not included in a training curriculum computerized.

Of these definitions, one sees that an educational object must make a comprehensive information and consistent. It can thus, as mentioned Yolaine Bourda [2], establish some characteristics of an educational object to know:

The autonomy
The reusability
The aggregation
The indexing

The term educational resource is used by Grandbastien [4]. This definition is very close of the definitions of the educational object.

The term component is used by Duitama [5] with similar definitions to the concept of educational object. The term component is generally used to designate everything that can return to the constitution of a training curriculum. It is therefore usually a catch-all term that we want to reserve to the concept of software component.

We propose the use of the term:

Educational resource for designate any numeric element containing a information, even partial, intended for teaching and likely to intervene in the constitution of a course, regardless of its physical size and its digital format. For example, an image, a schema or a demonstration of the theorems are educational resources. The "Assets" in the SCORM embody the concept that we join the term of educational resource. Everything that constitutes an interest in the reuse and can be used in isolation in training is an educational resource, because its recreating requires an investment on the part of users.

Educational Object to designate any unit of content consists of a set of objects and/or teaching resources. It has a pedagogical objective, an educational sense and a duration of use necessary to its assimilation. The SCO and the aggregates in the SCORM embody the concept that we join the term of educational object. It is equivalent to the English acronym LO (Learning Object). One or more learning objects can form a course and be reused in other courses. For example, a set of web pages with courses, exercises and demonstrations constitutes an educational object.

Educational component for designer any computer application can be integrated in a training curriculum and having disciplinary content. It is the case of some Java applets or software components that we find in warehouses of educational resources. In a training curriculum, a pedagogical component can be, according to the case, considered as an educational resource or educational object. The reverse is not always true. An educational object or an educational resource is considered a component pedagogical than if it includes instructions for treatment. Of this fact, a course HTML in line is not a pedagogical component while an applet of simulation of an experiment in physics is.

There is no border well defined between these three concepts. It must rely on the experience and the personal appreciation to determine whether an entity is a pedagogical resource or an educational object. By contrast, for whether it is a component of teaching, it is wise to ask the following question "the entity owns-t-it of processing instructions at the origin of a functional process executable on the processor?".

2 The Classes of Component Software

2.1 Educational Component Software

Component Educational Software (ECS) is a business component providing a pedagogical value. It participates in the process of human learning by ensuring educational features. Typical examples are the managers of pedagogical scenarios, simulators, the resolvers problems to pedagogical vocation or the tools for evaluation of the learner.

This notion of "business component" is consistent with the Approach software engineering led by MDA models (Model Driven Architecture) which advocates to separate the aspects the trades of technical aspects during the construction of an application.

From the point of view of software engineering, the class of the ECS is well a vertical layer containing components, qualified as vertical, dedicated to a type of specific applications and therefore necessarily used by a smaller number of developers. The class of ECS corresponds to the concept of component educational software used by Laforcade [6] and appointed by the Anglophones Educational Software Component - ESC.

2.2 Service Component Software

Service component software (SCS) is a business component providing a functional value devoid of pedagogy. It is helpful but not necessary, to ensure the pedagogical vocation of the platform. It offers a better ergonomics and manoeuvrability of the environment.

We classify in SCS two categories of components. First, the components belonging to other vertical layers (communication, medicine, military, statistics, etc.) that have been integrated in the platforms to provide services related to their field. These are some of the components belonging to other vertical layers parallel to the class of the ECS. Second, the components used by a restricted set of areas with needs convergent.

2.3 Technical Component Software

Technical component software (TCS) is a component providing the basic mechanisms necessary for the proper functioning of the CES and SCS. It is a component devoid of pedagogy and features of specialty. It is likely to be used in all areas.

In e-learning, they are mainly used for the construction of the infrastructure of platforms or as intermediaries between ECS and SCS little interoperable. This is the case, for example, of components of Persistence (backup of data), printing, formatting of text (transformation of an XML text in pdf), adapter of data structures playing the role of "glue" between components, etc.

3 Schema Dedicated to Software Component Metadata

In the face of the absence of metadata schema dedicated to components software and more specifically to those intended for teaching and learning, we have defined a schema to meet this need. For this, we have established a work plan with the following three phases:

A first phase during which we have, in a list, compiled and classified the elements turning around this problem in the models of existing metadata.

In a second phase, we have, but without reference to the previous study, prepared a list of useful metadata for the description of software components by adopting different points of view. (Designer, developer, assembler, user of software component) and by asking ourselves each time the following questions: "What would I like to know about this program or component? ", "What can I say about this program or component? ". The result was a list of various properties.

In a third phase, we have confronted these two lists. We have refined the organization of elements of our list, deleted the unnecessary elements and added the missing. This list has been validated by researchers of different disciplines (computer scientists, educationalists, educators).

The following diagram “Figure. 1” illustrates formally our remarks:

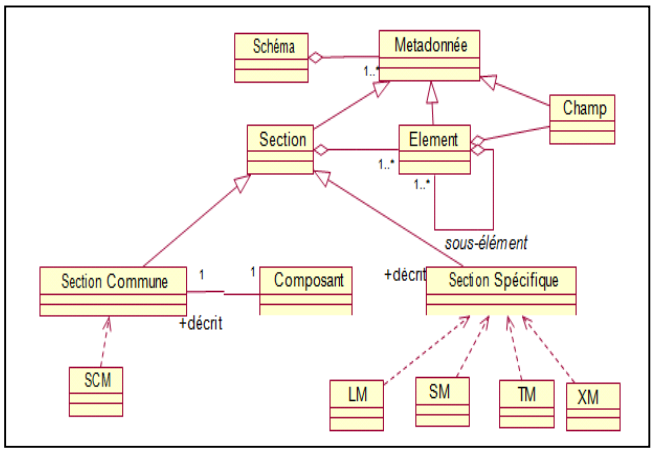


Figure1. Relationship component and sections of the diagram

Notes: The components not belonging to a domain of application can be described only with the software components Metadata (SCM) section. For the components belonging to several areas of application, we can adopt two strategies of Description:

Either, it is estimated that a domain is dominant in relation to the other, in terms of services offered and implemented by the component, and in this case, it is described with the metadata of the section specific to the field dominant.

Either one considers that the areas have the same level of dominance and in this case, it may describe the component with the sections specific to the areas dominant.

The choice of the strategy is left to the owner of the component. [7]

3.1 Result: The Metadata of the SCM

A schema is an entity composed of a set of metadata. At the internal level, each section contains a set of metadata hierarchically structured in the form of a tree. The nodes of the tree are called elements or sub-elements. They are used to organize the metadata. The leaves of the tree are called fields.

They are used to contain information. The sections, the elements and the fields are all metadata. These metadata are classified under a root element named SCM (see Fig. 2):

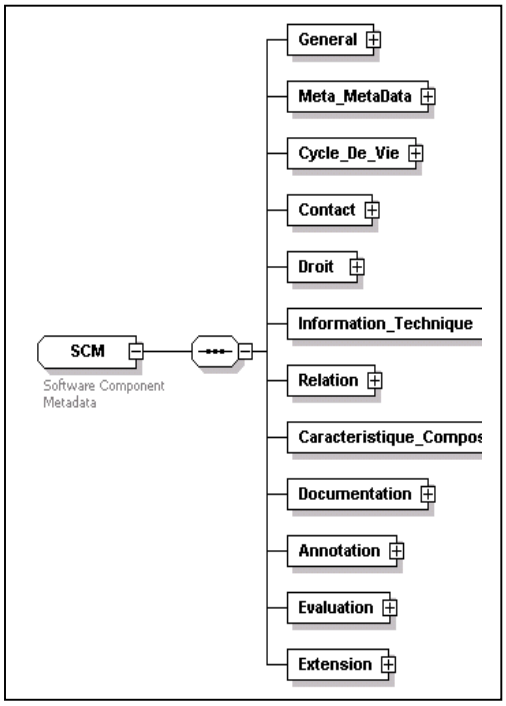


Figure2. The structure of the section "SCM"

The metadata of the SCM section allow you to describe all the software components independently of the field and the discipline. We have organized in twelve headings. These topics are for a share resumption of categories of metadata schemas as LOM and for another part of the specific additions to the needs of indexing, research and reuse of software component.

The metadata are used to find the components, they must be attached. That is why we propose a format of the package for the exchange and sharing of software components. We present, subsequently, the approach to follow to build this package and we will provide a prototype of an Intelligence Tool of metadata and generation of package.

4 Educational Component Repository

Before constructing the Educational Component Repository (ECR) commonly called warehouse, it is necessary to think about the features that users are entitled to expect from such a product and the best strategy to design in order to make a scalable, adaptable and generic [8].

4.1 The features of the ECR

The warehouse contains the software components developed by the Community e-learning. Each package of component is formed of a body and a description. The body is composed of binary code executable, the design model, source code, set of tests, documentation, etc. The description contains the metadata defined in the Fig. 3.

This fact the warehouse is composed of two parts: a catalog and a set of software components. The catalog includes the metadata of the components removed and especially information on their location. The components are archived in a file system.

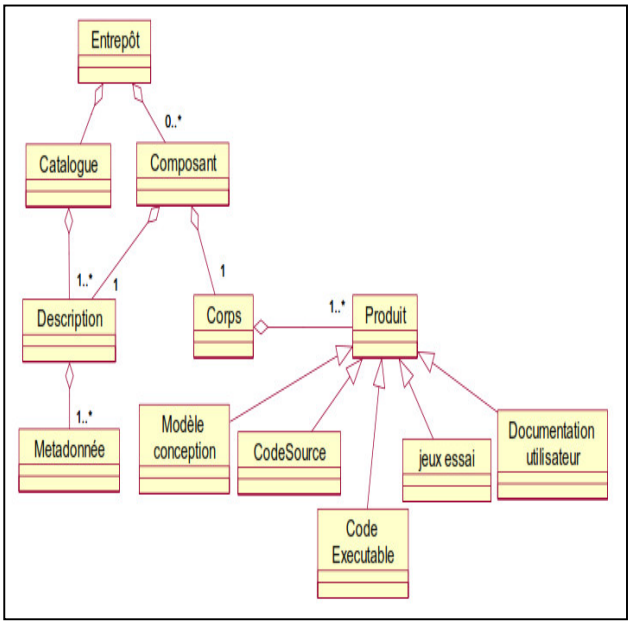


Figure3. Components of a warehouse

Identification and description of component software.

This information comes from the metadata of SCM and are recorded in the catalog.

Publication of software entities allowing developers to insert in the warehouse a new software component.

Route of the catalog allowing users to consult the description of the components software deposited in the warehouse.

Recovery of component software that allows users to obtain in their work environment a copy of the body of software component sought.

Classification and Search. In the case of a large warehouse, the course of sequential catalog is not enough. It must have the features of text search, by keyword, by attributes and by thesaurus.

For the manager of the warehouse the following features:

Storage and tracing information to use components.

Recording data on the number of access and retrieval of components, the number of successful research or not successful.

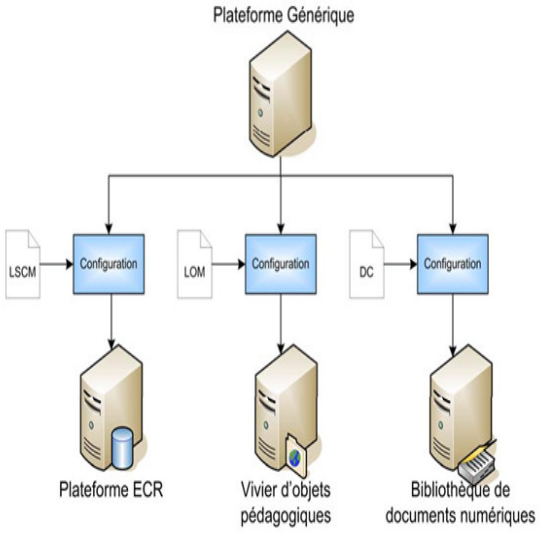
Control of access to the features offered and information contained in the warehouse.

Notification of changes to allow all potential user to be informed of the provision of a new component or an update.

4.2 The ECR an instance of a generic platform

The ECR should be scalable, adaptable and generic. These commitments has for consequence to complicate its design and realization because it must find a conceptual model of the platform allowing to have a independence between the metadata schema and the constituents of the warehouse: the model of the base of the index, the forms to fill used for the description and the indexing of software

components. In other words, the platform must be totally independent of the metadata that it will manage. It is in this sense that we are talking about generic platform.



Ensures the generic nature of the Platform implies that:

The basis of the index and the XML database are not formatted

The forms of interaction are not defined

The platform includes configuration tools in order to instantiate with the properties of the models of documents to be indexed.

4.3 Architecture of the ECR

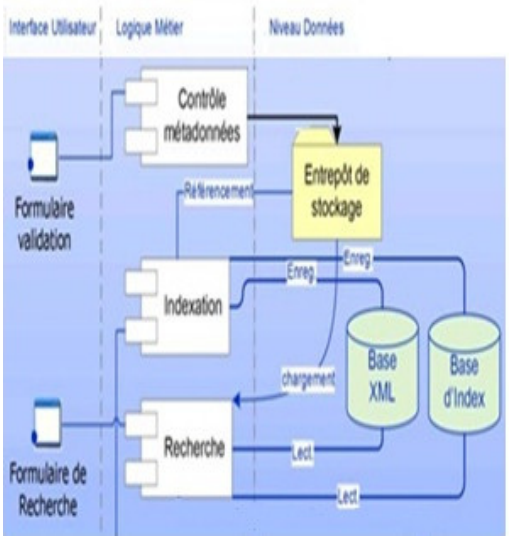


Figure5. Platform ECR (Educaional ComponentRepository)

The ECR is a web platform developed in the form of components, respecting the MVC architecture. The most important are the component configuration, the component of indexing and the component of research. The complexity of the system and the need to control the technologies used for its development do not allow to detail here the operation of its components. In this part, the presentation will be limited to the description of the broad lines of the process "index" and "Search". Fig. 5 shows the internal architecture of the platform and the connections linking its constituents.

4.3.1 Indexation of the components of the package

As soon as the transfer of the package in the storage warehouse, the metadata files are extracted and transmitted to the indexing tool. They are then stored in the XML database and indexed. The basis of the index is supplied by the result of the indexation. The package, him, is archived in a folder deposit. A message is then sent to the supplier to inform him that it is become the depositary of the package and its components. From this moment, the components can be searched via the search tool integrated into the platform.

4.3.2 Search for components in the warehouse

To perform a search of components, the user selects the search profile the more appropriate to these criteria. A research profile essentially defines the fields on which are carried out the research and the mode used. In the case of the ECR for example, it is possible to have a form of research intended to computer specialists, educationalists, teachers, etc. searches are carried out from the base of the index. On receipt of the search query, the system provides a list of components that meet the query criteria. The user then consults with the metadata of the components of interest. If it decides to retrieve a component, the package the container will be transmitted to him in full. The search in the database of the index is done according to two additional modes: the search by keyword (with the possibility of a multicriteria research) and the search by navigation.

In the search by keyword, we have similar features to the search engines on the Internet (genus Google). As far as we are concerned, we use internally the search engine full-text Lucene of Apache [9] which allows us to have a notion of relevance in the returned documents. The XML query language that we have created provides a result in an understandable syntax by Lucene.

In the search by navigation, there is the problem of the classification of the entity according to several criteria according to the expectations of the user. In effect, a complex entity, such as component software, is quite difficult to classify according to a single hierarchical tree because it can be viewed in different ways. To respond to this difficulty of classification of results (complex entities), we use the concept of classification by facet (Prieto-Diaz, 1991) [10]. This consists of a collection of "small" lists or hierarchies of orthogonal classification to each other.

5 Conclusion

This work has allowed us to see that designing a generic platform is a considerable work; both on the conceptual plan that on the plan of the computing development. For that this work has a chance to continue, it is essential that this platform is entirely independent of the diagram of metadata that it uses. "Entirely" means that all the useful data of the configuration (diagram of metadata, fields of research, structure of the basis of the index and the XML base) are external to the platform, all interfaces users (to

index and search for components) are automatically generated by the platform. As well, it is possible to produce different instantiations of the platform.

REFERENCES

[1] Bourda Y. and Hélier M. (2000). Métadonnées et XML : Applications aux "objets pédagogiques". Actes TICE 2000 - Technologies of information and communication in education for engineering and industry pages 135-141.

[2] Bourda Y. (2001). Objets pédagogiques, vous avez dit objets pédagogiques ? Cahier GUTenberg, n°39-40 pages 71-79.

[3] http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

[4] Grandbastien M. (2002). Quelques questions à propos de l'indexation et de la recherche de ressources pédagogiques sur le Web. Les technologies en éducation : Perspectives de recherche et questions vives pages 211-220.

[5] Duitama J.-F. (2005). Un modèle de composant éducatifs pour des cours adaptatifs sur le Web. France, Institut National des Télécommunication.

[6] Laforcade P. (2004). Méta-modélisation UML pour la conception et la mise en oeuvre de situations-problèmes coopératives, Université de Pau et des pays de l'Adour.

[7] AARAB Y, AKNIN N, and BENKADDOUR H (2016). Generation System of Metadata Software Component: A Proposed Architecture.Proceeding of Mediterranean Conference On Information & Communication Technologies 2015, Vol.381 of the series Lecture Notes in Electrical Engineering , pp 313-322.

[8] AARAB Y, AKNIN N, and BENKADDOUR H (2017).Component Based Implementation of Tool e-package. International Journal of Scientific & Engineering Research, Volume 8, Issue 1, January-2017 820 ISSN 2229-5518.

[9] Lucene, Apache Lucene project: http://lucene.apache.org

[10] [8] Prieto-Diaz R. (1991). Implementing Faceted Classification for Software

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

A Dynamic Clustering Approach for Maximizing Scalability in Wireless Sensor Networks

[1]Hanane Kalkha, [1,2]Hassan Satori, [1]Khalid Satori
[1]Department of Computer Science,
Faculty of sciences Dhar el Mahrez, USMBA
Fez, Morocco
[2]Department of Mathematics and Computer Science
Multidisciplinary Faculty, Mohamed First University
Nador, Morocco
hanane.kalkha@usmba.ac.ma; hsatori@yahoo.com

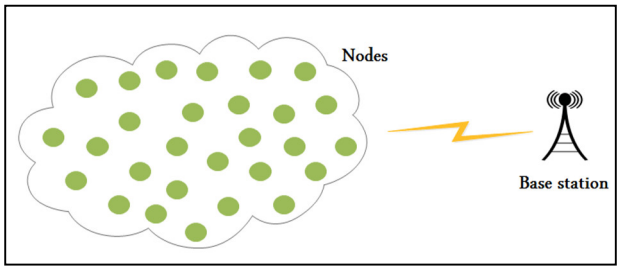
ABSTRACT

Scalability is an important and crucial issue which in routing protocols for Wireless Sensor Networks (WSNs). In this paper, we present an approach to achieving a balanced energy consumption rate using dynamic clustering to provide scalability in WSN. The proposed work in this paper is based on the dynamic clustering using k-means compared to LEACH (Low-Energy Adaptive Clustering Hierarchy), which is one of the most simple and effective clustering solutions widely deployed for WSN. The simulation results show that our proposed algorithm significantly improves high network scalability compared to LEACH.

Keywords- WSN; energy-efficient; K-means; LEACH; clustering;

1 Introduction

Wireless Sensor Network (WSN) is a popular research area, due to the potential use of sensor networks in different fields. A sensor network is a compound, of sensing, processing, communication ability to observe and react to events in a specified environment. WSN is usually composed of tens to thousands of nodes. These nodes collect, process and transmit cooperatively information to a central location. WSNs have specific problems such as a low duty cycle, power and life limited battery constraints, the acquisition of redundant data, heterogeneous sensor nodes, node mobility, and dynamic network topology, etc. [1, 2].



Fogure 1 Wireless sensor network architecture

DOI: 10.14738/tmlai.54.3328
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3328

WSN has interesting applications such as environmental monitoring and tracking applications, this was approved recently by the availability of smaller, cheaper, and intelligent sensors. These sensors have limited power resources generally irreplaceable. Therefore, when the ordinary networks aim to attain a high quality of service (QoS), WSN should focus first on the energy conservation. So, they must have integrated mechanisms that compromise offer the possibility to extend the network life at the cost of lower throughput and high transmission delay [3, 4].

Hence, the design energy-aware algorithms become an important factor to extend the sensor lifetime. Grouping nodes in clusters were widely followed by researchers to achieve the network scalability objective. Each group will have a chief often called the cluster head (CH) [5]. While many classification algorithms have been proposed in the literature for ad hoc networks. The majority of these algorithms is concerned mostly about the node accessibility and stability of the route without worrying about the critical design sensor networks [6]. Recently, several clustering algorithms have been specially adapted for sensor networks [7]. These available classification techniques varied widely according to the numbers of nodes, the network architecture, the characteristics of CH nodes and network process model. A CH can be elected by the sensors into a cluster or pre-allocated by the network designer and can also be simply one of the sensors, or a node which is higher in resources. The cluster composition can be fixed or variable. CHs can provide a second bearing network or may simply send the data to interested parts, for example, a base station or a command center [8].

The primary objective of hierarchical routing is to maintain effectively the energy consumption of sensor nodes by engaging them in multi-hop communication in a specific cluster and performing data aggregation and fusion in order to reduce the number of messages forwarded to the sink. Cluster formation is generally based on the energy reserved of the sensors and detection of cluster head [9]. LEACH protocol (Low Energy Adaptive Clustering Hierarchy) [10] is among the first hierarchical routing approaches for sensor networks. The proposed LEACH concept was an inspiration for many hierarchical routing protocols in the literature.

Even though the LEACH protocol act in a good manner, it also suffers from several disadvantages such: Firstly, The cluster head is generated randomly, so that cannot ensure the uniform distribution of CHs. This selection doesn't take into consideration transmission distance and the residual energy of the nodes, resulting uneven energy consumption for different nodes. Secondly, the high frequencies of re-clustering leads to a waste of energy [11].

The aim of this work is to propose an algorithm to extend the network lifetime. Our proposed algorithm was inspired from LEACH and based on k-means algorithm. The remainder of this paper is structured as follows: Section 2 presents the related work, Section 3 is about hierarchical routing and Section 4 discusses the scalability of WSNs, followed by description of used protocol in section 5. Section 6 present the simulation setup used. The results and discussions are presented in section 8 before concluding the paper.

2 Related work

Limited energy resources of sensor nodes create difficult issues on the improvement of routing protocols for WSN. Clustering is an important mechanism in wireless sensor networks to achieve scalability, reducing energy consumption and have a better network performance. Recently, H. Kalkha et al. [12] have compared in their work the performance of the two most famous routing protocols for wireless sensor

networks AODV and LEACH the first is plat based protocol and the second is a cluster-based protocol, the results showed that LEACH is more efficient than AODV for different parameters such as energy consumption, throughput and PDR.

Many research projects explored clustering in WSN from various perspectives. LEACH [10] is the first clustering algorithm has been proposed to reduce energy consumption. In which, the grouping task is set in rotation between the nodes, depending on duration. The data transmission between the CH and the base station (BS) is done through direct communication.

A modified release of LEACH known as LEACH-C [13] (Leach centralized) uses a centralized cluster formation algorithm to form clusters. Centralized approaches allow having an overall view of the network and therefore identifying the heads and members of the cluster. Even if all the information is available in this sort of approach, the determination of a proper value of the number of clusters is still a hard problem.

HEED (Hybrid, Energy-Efficient Distributed Clustering) [14] extends the basic of LEACH using the residual energy and the degree of node or density as a metric for cluster election in order to achieve the balancing power. It works in multi-hop networks by using adaptive power transmission in the inter-cluster communication.

PEGASIS (Power-Efficient Gathering in Sensor Information System) [15] inside instead of transmitting the packets of many cluster heads as such in LEACH, each node will form a chain structure to the base station through which the data is transmitted to the base station. PEGASIS achieves energy efficiency by forwarding data to one of its neighbor nodes.

3 Hierarchical Routing

Hierarchical or cluster-based routing are techniques that provide particular advantages of scalability and efficient communication. The hierarchical routing concept is also used to perform energy-efficient routing in WSNs. In a hierarchical architecture, higher energy nodes may be used to process and deliver the information while the lower energy nodes may be used to perform the detection in the proximity of the target. It means the creation of clusters and assigning special tasks to cluster head can contribute greatly to the overall scalability of the system, the lifetime and energy efficiency. So, hierarchical routing is an effective way to reduce energy consumption in a cluster and performing data aggregation and fusion, to reduce the number of messages transmitted to the sink [17]. In clustering approach, wireless sensor network can be considered of having following parts [18]: A group of nodes named cluster. Member nodes of the cluster. Cluster Head (CH) which is the leader of the cluster is charged for the delivery data from the cluster members to different cluster heads or base station Base Station (BS) which relay between network and end-user.

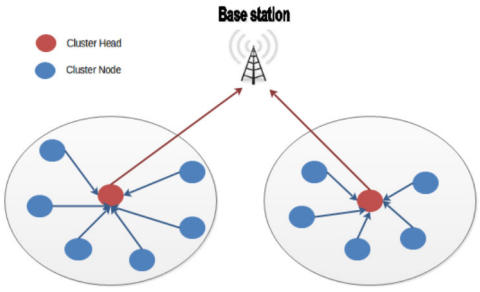


Figure 2 Schematic representation of clustering.

3.1 Classification of Clustering Schemes

Wireless sensor networks clustering schemes may be classified as distributed and centralized. In distributed approach, nodes locally exchange information for selection of CHs and formation of clusters. In a centralized approach, a central node, such as a base station control selection of CHs and formation of clusters [18].

For the state of clusters, clustering schemes can be categorized as dynamic clustering and static clustering. In the dynamic clustering, clusters are reformed after the end of the round. In the static grouping, clusters are formed in one and CHs are selected from cluster nodes.

According to the sensor node's characteristics, clustering algorithms can be classified as heterogeneous clustering and homogenous clustering. In Homogenous clustering the nodes are indistinguishable even there are super nodes. All nodes have equal opportunity to be CH. Heterogeneous clustering algorithms categorize nodes as normal nodes and super nodes. Super nodes are having higher energy than normal nodes and have a high chance to be selected as CH.

3.2 Dynamic Clustering

Dynamic clustering is widely used in hierarchical routing, the dynamic clustering technique aim to assign each pattern datum set for the cluster with the closest centroid. The dynamic clustering provides several benefits; For example, clusters are dynamically formed in response to particular triggered events. If a node detects an event with sufficient capacity, the node volunteers to act as a CH. The CH invites the other nodes to be as members of this cluster. The nodes may be selected on different clusters at different times with different roles [19].

4 Scalability of WSNs

Scalability is very important issue in the conception of routing protocols for sensor networks. A routing protocol classed as good and effective when it is adaptable to network topology changes. To determine the scalability in WSN, there are many parameters to consider as the number of nodes, node deployment, time, etc.) [20].

In order to evaluate the scalability of routing protocols, there are many indicators to analyze, such as the network lifetime, throughput and energy consumption, etc. [20]. Figure 3 shows the most important parameters considered in the WSN performance evaluation.

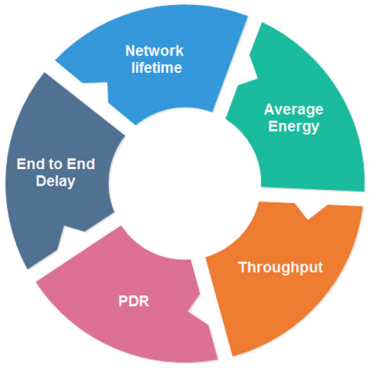


Figure 3 Scalability indicators

4.1 Lifetime Network

The most difficult challenges in the WSN design are how to conserve the node energy without affecting network operation. Moreover, the goal of an energy efficient technique is to maximize network lifetime, which strongly depends on the lifetime of every node.

There are several network lifetime definitions, depending on the wireless sensor network application. The network lifetime is defined in [21, 22] as the time until the first node/sensor in the network dies. In contrast, the network lifetime is defined in [10] as the time until all nodes die. The authors of [23] present the networks lifetime as the min{t1, t2, t3}, where t1 is the time it takes for the cardinality of the largest connected components to drop below c1 · n(t), where n(t) is the number of alive sensors at time t; t2 is the time it takes for n(t) to drop below c2 · n(0); and t3 is the time it takes for the area covered to drop below c3 · A, where A is the area covered by the initial deployment of the nodes. It is well-known that in both the ad-hoc and sensor networks, the network connectivity is very important to ensure the maximal delivery of the collected information [23].

4.2 Average energy

Energy consumption is the total energy consumed by the sensor nodes in the network [24], the energy consumption model based on the distance can be represented as follows:

$$E_{trans}(n,d) = \begin{cases} n * \left(E_{elect} + \varepsilon_{fmp} * d^2\right) & d < d_0 \\ n * \left(E_{elect} + \varepsilon_{amp} * d^2\right) & d \geq d_0 \end{cases} \quad (1)$$

$$E_{rec} = n * E_{elec} \quad (2)$$

Where,

- $E_{trans}(n,d)$:is the data transmit energy consumption,
- E_{rec}: is the data receiving energy consumption, n is the bit of the data packet.
- E_{elec}: is the wireless transceiver circuit energy consumption,
- ε_{fmp} and ε_{amp} : is the amplifier circuit power coefficient for free-space model and multi-path fading model respectively.
- d: is the distance between the sending node and the receiving node.

The figure below shows the energy model of L bits transmitted over the distance d:

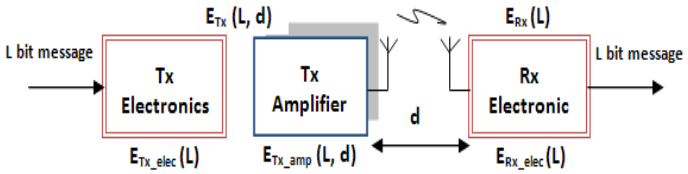


Figure 4 Energy Model in WSN

The equation used to compute average total energy (Eavg) per round is expressed as:

$$E_{avg} = \frac{Total\ energy\ consumed\ by\ all\ nodes}{number\ of\ nodes} \quad (3)$$

4.3 Throughput

Throughput is an important metric for improved network protocols, is usually measured in bits per second (bit/s) or sometimes in data packets per second (p/s) or data packets per time slot. The network throughput is the number of packets received successfully at the destination [25].

4.4 Packet Delivery Ratio (PDR)

The ratio of packets which are successfully delivered to a destination compared to the number of packets that were sent by the sender [25].

5 Description of protocols

5.1 LEACH protocol

Low Energy Adaptive Clustering Hierarchy (LEACH) proposed by W. Heinzelman et. al [10] is a clustering based protocol using the random rotation of local cluster to fairly distribute the energy among the sensors nodes. LEACH uses localized coordination to allow scalability and robustness for dynamic networks. LEACH integrates data fusion in the routing protocol in order to reduce the amount of information required to be sent to the base station. It reorganizes the network clustering dynamically and periodically, this makes it difficult to rely on a long lasting node to node confident link to make the protocol secure. LEACH assumes each node can directly achieve a base station transmitting with sufficiently high power. LEACH consists of two phases: Set-Up Phase and Steady State Phase Figure. 7.

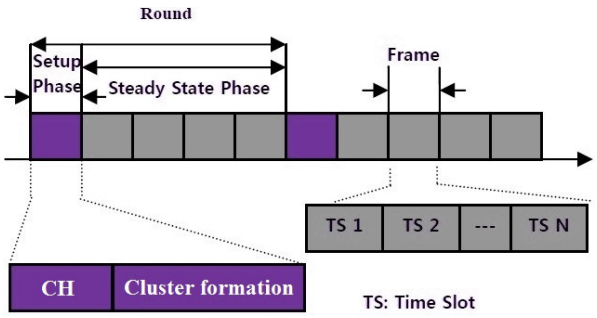


Figure 5 LEACH phases.

Set-Up Phase: At first each node computes a single random number between 0 and 1, and then calculates a threshold of formula T (n).

$$T(n) = \begin{cases} \frac{P}{1-P\times\left(r \bmod \frac{1}{P}\right)} & if\ (n\epsilon\ G) \\ 0 & else \end{cases} \quad (4)$$

With:

P: percentage of nodes desiring to become cluster-head.

r: current round number.

G: set of nodes that have not been elected cluster-heads during the 1/P last previous periods.

If the random number is less than the calculated threshold the node selected as a cluster head. After its selection each cluster head broadcasts an advertisement message to the rest nodes using the CSMA MAC protocol, and each node selects a cluster head based on the Received Signal Strength Indication (RSSI) of

the advertisement. Each node uses the CSMA MAC protocol to transmit its selection. After forming clusters, each cluster head create a TDMA schedule based on the number of nodes in the cluster.

Steady State Phase: Each sensor node collects and transmits data to its cluster head based on the TDMA schedule. Cluster heads receive all the data and send it to the base station. After transmission new round starts and executes the setup and steady state phase.

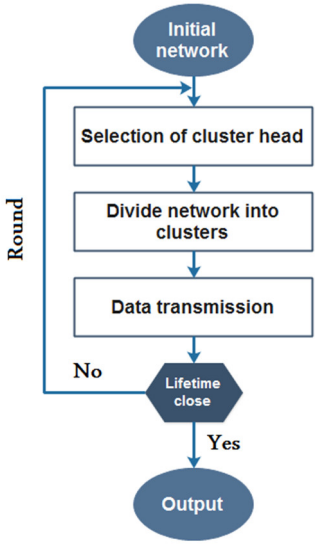


Figure 6 LEACH chart flow.

5.2 K-MEANS clustering

K-means algorithm is an unsupervised clustering algorithm, is a widely used technique for clustering based partitioning trying to find a number of specified number of clusters (k), which are represented by their centroid by minimizing the square error function developed for small size data often do not work well for high-dimensional data and the results may not be exact most of the time.

There are two simple approaches to initialize cluster center, i.e. either to choose the initial values randomly, or choose the first k samples of data points. Alternatively, different sets of initial values are selected (out of the data points) and the set, which is the closest to the optimal, is selected. [26].

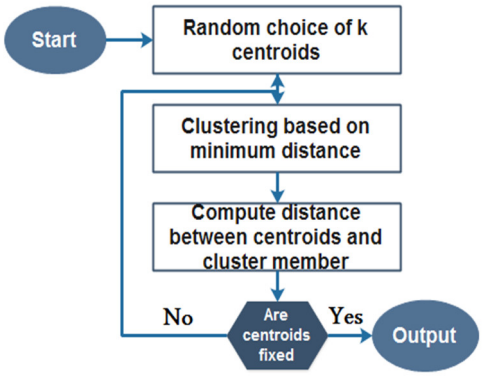


Figure 7 K-means clustering chart flow.

5.3 Our proposed algorithm based on K-means

Our proposed algorithm based on K-means is primarily use the Euclidean distances and selection of the cluster head node depends to residual energies. So, the central node collects the information about the node id, the position and the residual energy of all nodes and stores this information in the central node. After getting this information from all nodes, it starts to perform the clustering algorithm k-means.

Step of clustering

1. Choose k sensor nodes from all the n sensor nodes as the initial data centers.
2. Calculate the Intra-Cluster Distance [20] between other nodes to each of the centers using Euclidean distance as follows (1), and assign each node to the closest center.

$$\boldsymbol{intra} = \frac{1}{N}\sum_{i=1}^{k}\sum_{x\in C_i}\|\boldsymbol{x} - \boldsymbol{Z_i}\| \quad (5)$$

N is the number of nodes in the network,

K is the number of clusters,

Zi is the cluster center of cluster Ci.

3. Then assign other nodes to these clusters.
4. And calculate each node's cost. Node having the lowest cost is selected as cluster head.

This cycle continues until the K cluster heads fix down.

Data Transmission

Once the clusters are formed and the cluster head is selected.

1. The cluster head will specifies timeslot for each cluster member. Each cluster member transmits the data only in its timeslot**.**
2. The cluster head collects the data from the cluster member and forwards it to the sink node.

After data transmission, the network chooses the cluster head again.

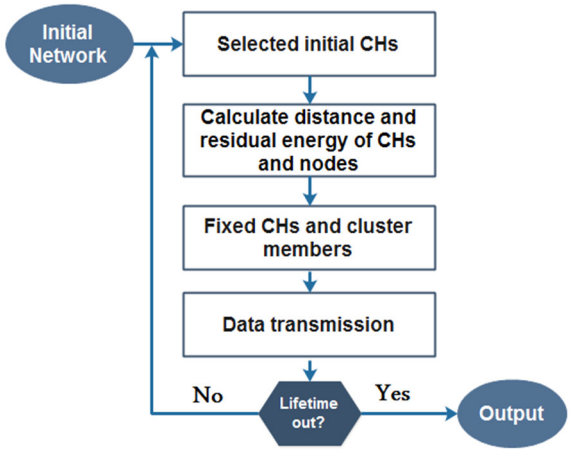


Figure 8 Our proposed clustering algorithm chart flow.

6 Simulation

6.1 Experimental Setup

In this work we use NS2 (Network Simulator 2) [30] wich is an object oriented, discrete event driven network simulator targeted at networking research. It provides support of TCP, routing and multicast protocol simulation on all wireless networks. NS2 can be employed in most UNIX systems and Windows. In this paper Ubuntu 14.10 LTS is used as operating system. Most of NS2 process procedure codes are written in C++ (The core of the simulator), TCL is used as scripting language and Otcl adds object orientation to TCL. The reason for using two programming languages is to have an easy, powerful and fast simulation.

The network model used in this work is as follows: All the sensor nodes in the network are homogeneous, not in mobility, uniformly deployed, and they have the same initial energy. The base station is fixed and located far from the sensor node. The number of cluster heads is 5% of the number of sensor nodes as in [31]. The rest of simulation parameters are shown in Table 1.

Table 1 Simulation parameters

Parameters	Value
Network size	250x250 m
Initial energy of node	100 J
TX, output power	-5 dBm
Eelec	50nJ/bit
Ɛfs	10 pJ/bit/m2
Ɛamp	0,0013 pJ/bit/m4
EDA	5 nJ/bit
Packet length (number of bits in the packet sent from CH to BS)	6400 Bits
MAC type	Mac/802_11

7 Simulation results and analysis

We evaluate the performance of the both protocols LEACH and our proposed algorithm based on K-means using the setup cited above. Results were obtained by means of 10 simulation runs and calculated the average of the measurements.

7.1 Time taken to cluster

Both algorithms have the same configuration and the positions of nodes are constant. The time taken is independent of positions of nodes, but it depends only by the variation of nodes numbers and position of cluster head.

We measure the time taken to cluster including time taken to exchanging adverting messages, e.g. time taken to exchanging the position and energy details with all nodes and clustering time and time taken for computing algorithm. In the case of LEACH the time taken to cluster includes also the time taken for the nodes to send back its selection to cluster head.

Figure 9 Time taken to form clusters

Analyzing figure 9 we can see that the time taken for LEACH clustering is longer than the time of our proposed clustering algorithm. Probably this is due to the time taken by nodes to send back its selection to cluster head in LEACH algorithm.

7.2 The average energy

In general, for clustering, energy is consumed principally for transmitting, receiving packets and for processing. The energy consumed by varying the number of nodes is shown in figure 10.

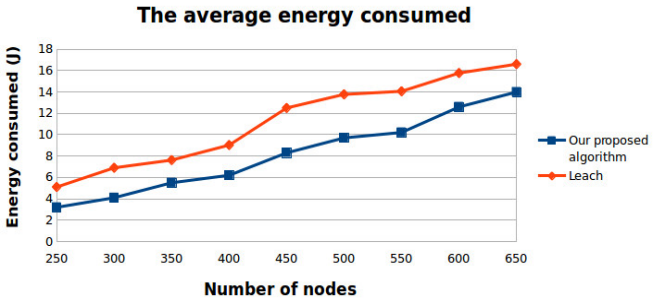


Figure 10 The average energy consumed per node

The average energy consumed per node is measured by computing the difference between the total initial energies of all nodes and total residual energies of all nodes after clustering dividing by the total number of nodes.

Table 2 Energy consumption reduction

Number of nodes	Energy consumption reduction (%)
250	37.25
300	40.75
350	27.82
400	31.33
450	33.6
500	29.55
550	30.06
600	20.01
650	18.37

Table 2 shows that our proposed algorithm based on k-means reduces the energy consumed compared to LEACH. The percentage of energy consumption reduction is between 19 % and 40 %.

7.3 Network liftime

The network lifetime is the time until the fraction of alive nodes falls below a predefined threshold or the time during which at least k out of n nodes are alive.

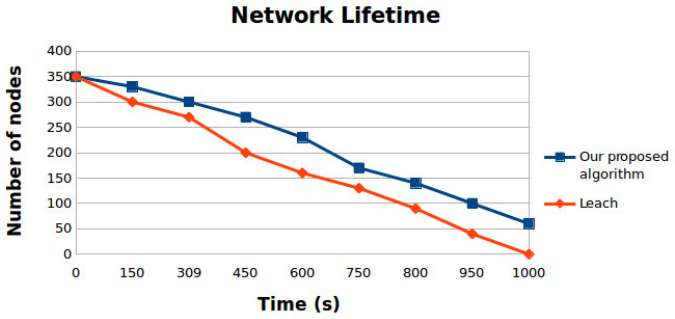


Figure 10 Network lifetime.

As shown in the results of the figure 11 we find that our proposed algorithm based on K-means is more efficient in terms of network lifetime compared to LEACH. Table III illustrates the network lifetime prolonging percentage when using K-means compared to LEACH routing protocol.

Table 3 Nodes Lifetime Prolonging

Time(s)	Nodes lifetime prolonging (%)
150	10
300	11
450	35
600	43,37
750	30,76
900	55,55
1050	66,66

7.4 Throughput

The throughput is depending on the number of packets received by BS.

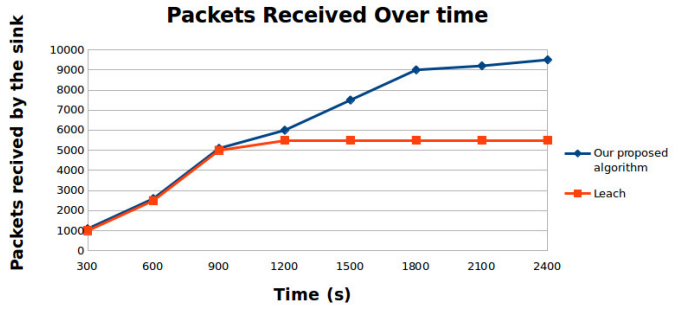


Figure 11 Throughput of LEACH vs our proposed algorithm

We can illustrate from the Figure 12 that the packets received by the BS using our proposed algorithm based on K-means, are considerably superior to LEACH. This improvement can be justified by the fact that

when using K-means network lifetime is more extended. Table IV shows the Throughput increasing percentage when using our proposed algorithm compared to LEACH routing protocol.

Table 4 The Throughput Increasing In Case Of K-Means

Time (s)	Nodes lifetime prolonging (%)
300	10
600	4
900	7,84
1200	9,09
1500	36,36
1800	70
2100	67,27
2400	72,72

From our simulations, we summarize the obtained results as follows: The throughput received and network life ameliorated, respectably, by 34.66% and 36.04% in the case of our proposed algorithm compared to LEACH protocol. Moreover, the average energy consumption decreased by 29.85%.

8 Conclusion

Our studies were focused on performance analysis of algorithm based on K-means in WSN. Several scenarios were carried out using different network settings. Our proposed algorithm was compared to a classical algorithm LEACH using the NS2 simulator. The results show that the proposed algorithm offers a considerable energy consumption reduction and extends the life of the network is remarkable and that K-means have the capability to improve the best packet ratio sending by the sensor nodes to the base station. In view of this study, we proposed to extend this work to examine other parameters such as robustness, reachability and QoS, etc. Also, it would be interesting to examine these parameters with other clustering algorithms and to show a comparative study of different clustering methods combined with traditional routing protocols.

REFERENCES

[1] A. ABBASI and M. YOUNIS, "A survey on clustering algorithms for wireless sensor networks," Computer communications, vol. 30, no 14, pp. 2826–2841, 2007.

[2] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Computer Networks, vol 38, no 4, pp. 393–422, 2002.

[3] Khan, T. Israfil, A. Emdad and A. A. Muhammad, "Multiple Parameter Based Clustering (MPC): Prospective Analysis for Effective Clustering in Wireless Sensor Network (WSN) Using K-Means Algorithm," Wireless Sensor Network, pp. 18–24, no 4, 2012.

[4] M. N. JOSHI, "Parallel k-means algorithm on distributed memory multiprocessors," Computer, vol. 9, 2003.

[5] K. Akkaya, and Y. Mohamed, "A survey on routing protocols for wireless sensor networks," Ad Hoc Networks, no 3 , pp. 325–349, 2005.

[6] L, Xu, O. GMP and R Collier, "A Balanced Energy-Efficient Multihop Clustering Scheme for WSN," Wireless and Mobile Networking Conference WMNC, pp. 1–8, May 2014.

[7] X. Liu, "A survey on clustering routing protocols in wireless sensor networks," Sensors , vol. 12, no 8, pp. 11113–11153, 2012.

[8] G. Gupta and M. Younis, "Load-Balanced Clustering in Wireless Sensor Networks," Proceedings of the International Conference on Communication, vol. 3, pp. 1848–1852, Anchorage, Alaska, May 2003.

[9] A. ABABNEH and E. AL-ZBOUN, "EDAC: A Novel Energy-Aware Clustering Algorithm for Wireless Sensor Networks, " International Journal of Advanced Computer Science & Applications, vol. 1, no 7, pp. 333–338, 2016.

[10] W, Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless sensor," Proceeding of the Hawaii International,vol. 2, pp.10, Hawaii, January 2000.

[11] R. Kaur, D. Sharma and N. Kaur, "Comparative Analysis Of Leach And Its Descendant Protocols In Wireless Sensor Network," International Journal of P2P Network Trends and Technology, vol. 3, Issue 1, pp. 51–55, 2013

[12] H. Kalkha, H. Satori and K. Satori, "Performance Analysis of AODV and LEACH Routing Protocol," The International Journal of Multi-disciplinary Sciences - ISSN: 2421-9606, Issue 2 , vol. 2 , pp. 56–60, February - March 2016.

[13] N. Nasri, A. WEI, A. Kachouri and S. E. Khediri, "A New Approach for Clustering in Wireless Sensors Networks," International Workshop on Wireless Networks and Energy Saving Techniques (WNTEST), vol. 32, pp. 1180–1185, Elsevier, 2014.

[14] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," IEEE Transactions on Mobile Computing, vol.3, no 4, pp. 366–379, 2004.

[15] S. Lindsey and C. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems," IEEE Aerospace Conf. Proc 3, pp. 3-1125 - 3-1130, 2002.

[16] K. P. Noufal, " Wireless Sensor Networks – Scalability and Performance Issues: A Review," IJCST, vol. 6, Issue 1, pp. 139–140, Jan - March 2015.

[17] J. N. AL-KARAKI and A. E. KAMAL, "Routing techniques in wireless sensor networks: a survey," IEEE wireless communications, vol. 11, no 6, pp. 6–28.

[18] P. Vipin, S. Girdhari and R. P. YADAV, "Energy efficient clustering scheme for wireless sensor networks: a survey," Journal of Wireless Networking and Communications, vol. 2, no 6, pp. 168–174, 2012.

[19] S. GANESH and R. AMUTHA, "Efficient and secure routing protocol for wireless sensor networks through SNR based dynamic clustering mechanisms," Journal of Communications and Networks, vol. 15, no 4, pp. 422–429, 2013.

[20] L. K. Alazzawi, A. M. Elkateeb, A. Ramesh, and W. Aljuhar, "Scalability Analysis for Wireless Sensor Networks Routing Protocols," 22nd International Conference on Advanced Information Networking and Applications, IEEE, pp. 139–144, 2008.

[21] J. H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad hoc networks", Proc. IEEE INFOCOM, pp. 22–31, Mar. 2000.

[22] K. Toh, "Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks", IEEE Communications Magazine, pp. 138–147, June 2001.

[23] M. Blough and P. Santi, "Investigating upper bounds on network lifetime extension for cell-based energy conservation techniques in stationary ad hoc networks," Proceedings of the 8th annual international conference on Mobile computing and networking, ACM, pp. 183–192, 2002.

[24] N. A. Pantazis and S. A. Nikolidakis, "Vergados, Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 2, pp. 551–591, 2013.

[25] L. ALAZZAWI and ELKATEEB, "Performance evaluation of the WSN routing protocols scalability," Journal of Computer Systems, Networks, and Communications, 2009.

[26] P. Sasikumar, S. Khara, "k-MEANS Clustering in Wireless Sensor Networks", Fourth International Conference on Computational Intelligence and Communication Networks, IEEE, pp. 140–144, 2012.

[27] V. Kawadia, P. R. Kumar, "Power Control and Clustering in Ad Hoc Networks," Proceedings of IEEE INFOCOM. San Francisco, CA, pp. 459–469, March 2003.

[28] J. YICK, B. MUKHERJEE, and D. GHOSAL, "Wireless sensor network survey," Computer networks 52, no. 12, pp. 2292–2330, 2008.

[29] F. ZHAO, Y. XU and R. LI., "Improved LEACH Routing Communication Protocol for a Wireless Sensor Network," International Journal of Distributed Sensor Networks ,2012.

[30] Network Simulator 2: Version 3.35, http://www.isi.edu/nsnam/ns/,, Accessed January 2016.

[31] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Computer networks, vol. 38, no. 4, pp. 393–422, 2002.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Towards a Model of a Scalable Middleware Architecture Based on Cloud Computing, Application Reliable Integration and E-Learning Platforms in Academic Information Systems

Abdellah Lalaoui Hassani, Khalifa Mansouri
Laboratory: Signals, Distributed Systems and Artificial Intelligence (SSDIA)
ENSET Mohammedia, University Hassan II OF
Casablanca, Morocco
lalaoui_abdellah@yahoo.fr; khmansouri@hotmail.com

ABSTRACT

Throughout the present article , we will examine, on the one hand, systems integration approaches of e-Learning applications in university's information systems. On the other hand, we will establish a study about the integration of applications between the internal information systems and the Cloud. Then between the applications on the private cloud and the public cloud.

Keywords—E-Learning; Higher Education; University's Information System; Middleware Architecture; Integration; EAI; SOA; ESB; Cloud computing.

1 Introduction

For more than a decade, the world of information systems has gradually invaded the Moroccan higher education sector. Universities, engineering schools, public and private higher education schools, are making great strides for an optimal use of information and communication technologies (ICT).

Several institutions have urbanized their information systems by either integrating management softwares such as Enterprise Resource Planning (ERP) [16] or locally developing systems, that are basically dedicated to manage:

- School affairs: notes, absences, timetables, agendas, etc;
- Internal and external communication: messaging, staff's information;
- Educational means/tools: digital textbook, document storage, collaborative tools, blogs, forum etc.

Those information systems are most or time developed to be accessible from any device (desktop, laptop, tablet, smart-phone) connected to the Internet. There are several software tools of managing school governance and school life management, for example :

The APOGEE system as a standard, is a paying ERP and the most used by universities. It meet, practically the majority of their own needs [6];

The GESFO system in particular, is a ERP developed locally by the IT Department of ISGA (Higher Institute of Engineering and Business) [10].

Figure 1 shows the general layout of the IS architecture within a Higher Education institution :

DOI: 10.14738/tmlai.54.3329
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3329

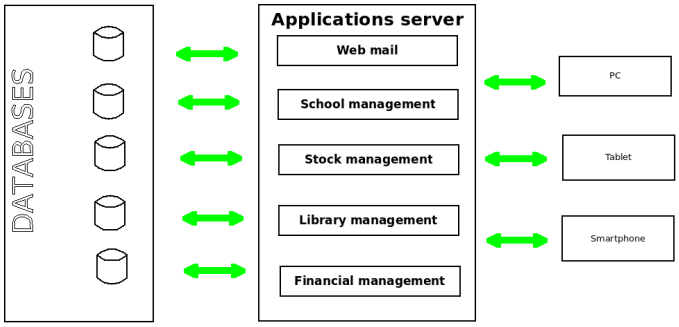


Figure. 1 IS architecture within an institution

ICT has greatly infested the world of pedagogy in Morocco [6], so to offer a large-scale interaction participation and an efficient control of its different users. for example, distance education, trough the use of e-learning

systems [7]. Those systems represent, first of all, a new teaching method that saves a lot of time, shares knowledge, builds skills and know-how. Secondly, they improve learning services for learners [8].

In addition, the Moroccan Ministry of Higher Education, Training of Managers and Scientific Research has shown a strong desire to successfully integrate e-Learning in education. Indeed, the managers are aware that this integration will surely improve the quality of teaching and learning. For this purpose, it is necessary that some Moroccan institutions aligns on international standards in terms of integration and use of e-Learning platform officially in their educational system. However, it is a large scale project, whose encounters so many difficulties and obstacles.

This article presents two studies:

The first one is about applications integration's field in a university's information system and its relationships with an e-learning platform.

The second one is related to the integration of applications between the internal information system and the Cloud (scenario IS Campus to Cloud and Cloud to IS Campus) then between applications on the Cloud (scenario from private cloud to public cloud).

After an introduction, which basically aimed at a contextualization of the concept of integration, we will present in a first section an overview about the different platforms of application integration, starting from the point-to-point architecture to the ESB (Enterprise Service Bus) architecture [3], by passing through EAI architecture (Enterprise Application Integration) [1]. Then, we will identify highlights of the ESB architecture.

We shall present, in the second section of the article a modern and a service-oriented architecture (SOA) system's structure and integration infrastructure in between two internal information systems (IS2IS)[3] and that, by using an enterprise service bus (ESB). Its aim is to create more flexible solutions for integration and coupling of systems. The third section of the article displays a presentation of cloud computing and as-well-as its main deployment model.

Thereafter, we will expose the advantages of e-learning based on the Cloud and its differences compared to e-learning based on a basic infrastructure. The last section presents a new model that is more relevant and beneficial for universities. This model combines the ESB integration solution and the Cloud computing.

2 Context of Our Work and Problematic Proposal

Today, we cannot imagine a software that can fit all the needs expressed by a company. An Information System (IS) is an assembly of independent applications, each of which is designed to meet a specific need. In order for the IT department to suit the requirements of the company, it tries to take the best customer relationship management (CRM)[16] solution, the best Enterprise Resource Planning Software (ERP) and the best Document Electronic Management solution (EDM)[9], so that in order to integrates them into its IS.

That is, as well, a particular problematic in the Higher Education sector. Following a study carried out on Moroccan institutions organized on several campuses, it was found that most of them gather several softwares (paying or open source) by means of heterogeneous solutions to adapt their University's Information Systems (UIS), with their specific learning management needs.

This configuration of the UIS causes a desynchronization of all the data between the applications. This will force users to duplicate information many times in different applications, and this could be a source of many input errors.

We have taken for this research, an institution of Private Higher Education in Morocco, composed of five training campuses and a general direction, spread over the national territory. In each campus, the information system is represented as follows:

Educ server: this server hosts an application named Educ that allows to manage all the business process of the school (students, absences, notes, teachers, etc.).

Educ DIRECT Server: This server hosts an application dedicated to tutors, in order to monitor the academic progression of the students. The data is replicated in real time between the two applications through a tool, which allows to automate the copying of the data between the Educ server and the EducDIRECT server in a unidirectional way. Each IS Campus works in an autonomous manner and there is no connection between them.

In 2015, the General Direction decided to integrate an e-Learning system as a complementary support to classical instruction called "face-to-face" (requiring the student's daily presence in the classroom). It chose the "Moodle" platform as an e-learning solution for training and distance learning. This platform operates independently of the other campus ISs.

It is hosted in the information system of the General Direction and regroup all students and teachers of the various campuses. With this solution, teachers can upload their courses in the platform and learners can take courses distantly.

Figure 2 shows the general layout of the overall architecture of the new IS within the institution studied by integrating the e-Learning platform.

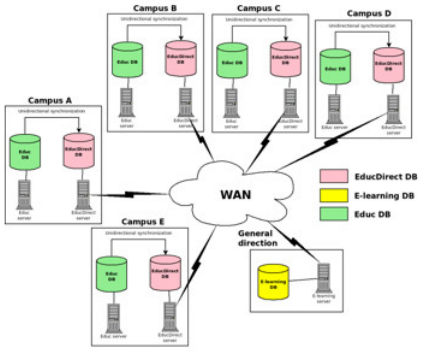


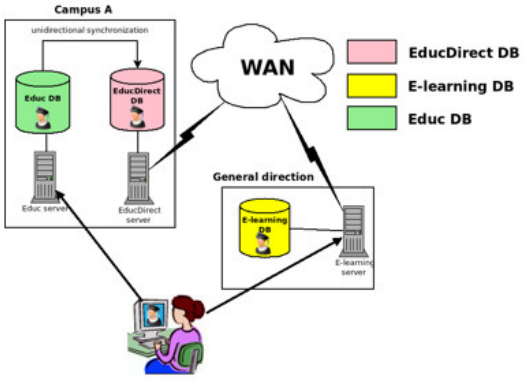
Figure. 2 Overall IS Architecture of the Institution Integrating the e-Learning Platform

This solution has raised a lot of problems at the business processes level of the campuses, particularly for the actors (students, teachers and operator of the studies department). These problems are:

Problem 1: Students and teachers have an access code for each application; therefore numerous access codes may be forgotten by their users.

Problem 2: The independence of this platform with ISs of the different campuses generates a delay to provide a synchronization of student and teacher accounts.

Problem 3: Operators are obliged to enter the same data twice (student information, notes, modules, user accounts, etc.) into each application, which requires a considerable time to feed the databases. Also, it may generate numerous errors from the data entered. The next figure presents an illustration of this problem:



Figiure. 3 Illustration of problem 3, posed by the global architecture of IS.

Figure 3 shows that the operator connects to the Educ application in order to enter a new data of a student, then it must switch to the e-Learning system to re-enter the same information of that student.

3 Proposal for a Technical Solution

In order to provide a better user experience for learners, educators and managers and to avoid the incoherence in the data because of actions that are performed on a system but not on another, or to enter the same data in a variety of systems, it is necessary to ensure integration of information exchange between all these applications.

Our approach is to propose an integration application platform. This platform allows heterogeneous applications to manage their exchanges while ensuring the confidentiality, integrity, speed of transfer and authenticity of the data. For instance, it allows an application, viewing and / or updating data from another application.

Figure 4 illustrates the different exchanges of data allowed between the e-Learning platform hosted in the information system of the general direction from any campus.

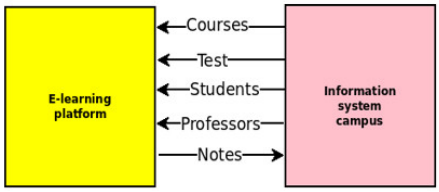


Figure. 4 Exchanges of data between the e-Learning platform and the IS campus

The integration of applications is an issue that was born with the evolution of the architectural and technological level of the software. According to Manouvrier, integration brings all the methods and tools that organize exchanges between the applications and the business processes of the company[1].

3.1 Forms of integration

Three forms of integration can be distinguished:

Integration of data;

Integration of specific treatments;

Integration of global applications.

Data integration is a form of integration dedicated to databases in order to facilitate their connectivity. This form allows the user to migrate, consolidate, combine all types of data from various sources: flat files, database, mail, etc.

Process integration is a form of integration that consist on integrating a new software entity within an application to provide an additional functionality, such as Application Programming Interface (API)[3].

Application integration consists on an approach that allows

Interconnecting distributed and heterogeneous applications developed independently in incompatible way, such as the case of two ERPs coming from two different editors.

Application Integration Platforms

Three approaches of architectures integration can be identified:

Point-to-point architecture;

EAI architecture;

ESB architecture.

3.2 Point-to-point architecture

It is an architecture based on a point-to-point integration approach. Each pair of applications is considered as separate brick, related to its own connectors and its own communication protocol (Figure 5).

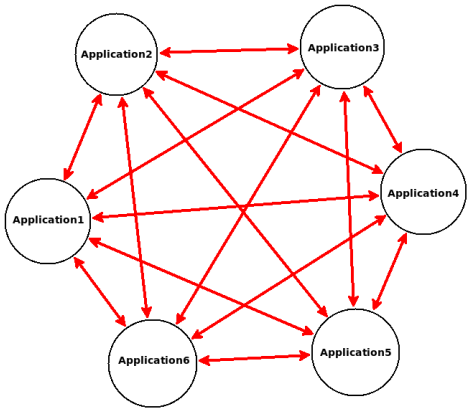


Figure. 5 Point-to-point integration architecture

This type of integration causes a confusion of proprietary and heterogeneous connections, which makes it intricate for application to be managed since the number of interfaces increases with the number of applications that need to be integrated into the information system.

3.3 EAI Architecture

This architecture is a centralized platform. Each application connects in a unique and independent way to the central system, without knowing a priori, the global topology of the information system (Figure 6).

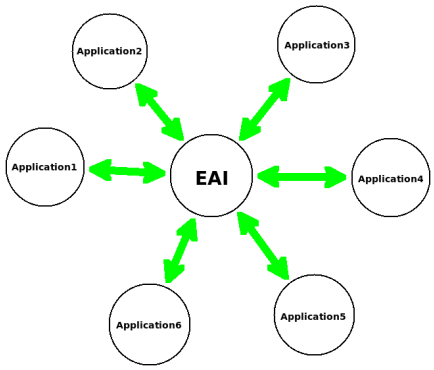


Figure. 6 EAI Integration Architecture

Despite its advantages, compared to the first architecture, in terms of robustness, reliability unification of connectors, and as in any centralized architecture, the main weakness of this architecture is the SPOF. (Single Point of Failure). Indeed any malfunction at the central system, leads to shutdown the entire platform.

3.4 Architecture ESB

It is a fully distributed architecture, often used in the SOA context [11] and providing services such as data transmission, data transformation and increased interoperability through the systematic use of standards such as XML (Extensible Markup Language)[14] and WEB SERVICES [12], (Figure 7).

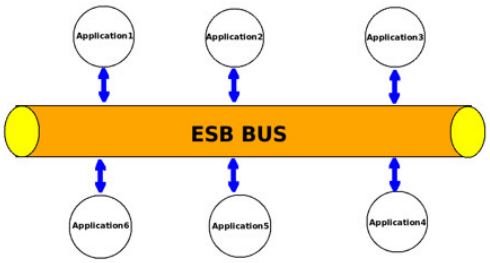


Figure. 7 ESB integration architecture

On the basis of this observation, and after comparing the different approaches mentioned above, we have opted to use the ESB integration architecture as an appropriate solution for our case study [11].

4 Initial Implementation of the Esb Architecture in Our Case Study

Figure 8 illustrates the implementation of the ESB Integration Architecture on a school campus.

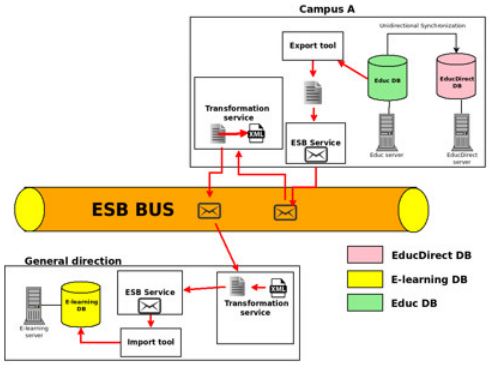


Figure. 8 Implementing the ESB Integration Architecture on Campus A

The data passing through the ESB bus is always represented by the file generated by the export tool. The transformation of the representation of this data in an XML form will allow to share this information on the bus.

The structure of the file generated by the export tool is understandable only by the import tool, located at the seat.

Data transformation takes place in the ESB bus. A transformation service is intercalated in the process of sending messages to the headquarters. This service is configured first, to recognize, the structure of the file generated by the export tool, second, to generate a better structured file.

The data transiting the ESB bus is not sent directly to the recipient, but will follow a circuit passing through different points of the ESB until finally reach the seat.. This circuit is defined directly in the message containing the data.

Returning to the problem number 3 that has been already explained in figure 3 and after the implementation of the ESB integration architecture, the campus A operator will enter data that concerns one student, only once on the Educ system. The data will pass through the ESB bus, as a generated file throughout the export tool. This file will be transformed into an XML form, through the transformation service. Once the XML file arrives to the general direction, The transformation service will convert this file into an understandable one by the import tool. Then this tool will send the file that contains the information of a student to the E-Learning database.

In the case of a technical problem that requires the restart of the bus or when a recipient (here, the general direction) is not reachable momentarily, after that the e-Learning server will shutdown, the non-delivered messages will be sent again to a backup system integrated in the bus [3].

5 Improving the Proposed Solution

The proposed architecture has considerably improved the business process of the institution studied, namely:

- Avoid to re-enter notes, student information, modules and matters;
- Avoid to recreate accounts of students and teachers;
- Avoid to give two different accounts to students and teachers;
- Adapting the mixed learning method with the information system such as course follow-up, inputting notes etc.

However, the data transmitted on the ESB bus is sometimes voluminous, which involve that the data transformation tool can take a long time in processing before transferring it to the recipient.

We have tried, through a new approach, to improve the transfer time of data sent by all applications.

To do this, we focused on two main points:

The nature of the data transferred;

The data transfer processing on the ESB bus.

5.1 Nature of data transferred

Applications that exist on campus and seat can generate data in XML format, Then we can add a coding tool in front of each application that will encode the data generated in XML to EXI and reciprocally EXI to XML.

EXI is a binary XML storage and data transfer format that has the same mechanism as the XML format, but is more compact, faster to exchange and process on the computer network [2].

5.2 Data transfer processing on the ESB bus

To reduce data transfer processing, we have moved the ESB bus directly in front of each encoding tool without having to connect with the export tool.

A JCA (Java Connector Architecture) adapter is used to connect the applications to the ESB.

This adapter allows sending information about, notes, students or teachers on the ESB bus.

This information will be directly EXI elements of type note, student, teacher, module, subject as already seen. The same for the received data, they contain information in EXI format that will be interpreted and transmitted to the destination application (Figure 9).

Figure 9 illustrates the new version of the ESB architecture with improvements in the nature of the transferred data and the processing of data transfer on the ESB bus.

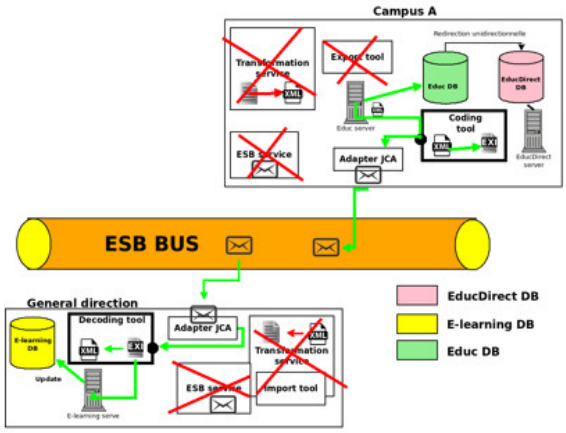


Figure. 9 The final version of the proposed ESB architecture

This change in the architecture of the institution's information system ensures the following advantages:

- The exchange of information between the various campuses and the seat is instantaneous and rational;
- Data is secured across the bus;
- Data transfer is fast through the bus;
- The information is represented in standardized form, and can be read and used from any application of the establishment;
- The information is entered once by the operators of the system.

Despite its technical advantages, this traditional online learning mode illustrates many disadvantages that can be presented in the following way:

- The development and maintenance of the systems is done by the educational institutions, which requires a permanent and competent IT team to ensure the evolution and high availability of the system;
- In order to ensure the well functioning of the e-Learning platform, the institutions must pay for software licenses, technical support for software and also verify that the software is installed on client devices;
- This architecture has no backup and restore policy, which can cause a big problem if the server crashes;
- Lack of an online collaboration and storage space for teachers and students which obliges them to back up and transfer files from one device to another.

6 Proposal for a Technical Solution Based on the Cloud Computing

Since the traditional e-learning solution is confronted with several problems related to resource optimization, storage, simultaneous data processing and also increasing costs, it becomes necessary to use an environment that meets both, scalable requirements and cost control. This environment is named the Cloud computing.

Our approach is to propose a modern, service-oriented infrastructure while maintaining the ESB middleware architecture that will ensure the integration of applications between firstly, two internal information systems (Scenario between a campus's internal IS and the e-Learning platform hosted in the general direction). Secondly, between the internal information system and the Cloud (Scenario between a campus's internal IS and the e-Learning platform hosted in the cloud) while ensuring high availability , Data security and IS to Cloud, Cloud to IS and Cloud to Cloud communication.

6.1 The Cloud computing

According to NIST [13], cloud computing is a model of a service platform that allows easy, on-demand access by the network to a shared set of computing resources (servers, storage, applications, and services).

Cloud computing consists of four deployment models:

Private cloud: used by a single organization, it can be managed by the organization or by a third party. It can be placed in the organization's locals or outside.

Public cloud: open to the public. It is owned by an organization that sells Cloud services.

Community Cloud: shared by several organizations for the needs of a community that wish to pool resources (security, compliance, etc.)

Hybrid Cloud : composed of one or more models (public, private and community). These models are linked together by the same technology that allows portability of applications and data It is an excellent solution that allows to benefit from the advantages of private clouds and public clouds [15].

The National Institute for Standardization and Technology (NIST) divides cloud computing into two types of models:

Service models that refer to particular services accessible throughout the cloud computing platforms,

Deployment models that refer to the location and management of the Cloud infrastructure.

6.2 E-Learning based on cloud computing

With the increase of student's number, the rapid growth of educational content and the evolution of the IT infrastructure, educational institutions are facing the increasing costs and the decreasing budgets. Which requires to find an alternative for their e-Learning solutions.

Furthermore, current e-learning systems are not scalable and do not lead to efficient use of resources. In response to these barriers, institutions need to adopt service-oriented approaches.

The potential effectiveness of using cloud computing in higher education has been recognized by many international universities, such as the University of California, Washington State University School of

Electrical Engineering and Computer Science, Higher Education Institutes of United Kingdom, the United States and others [12].

Institutions will be responsible of the process of education, content management and delivery while the service provider is responsible of the development, management and maintenance of the system. Establishments are charged according to the number of servers used that as well depends to the number of students as shown in figure 10.

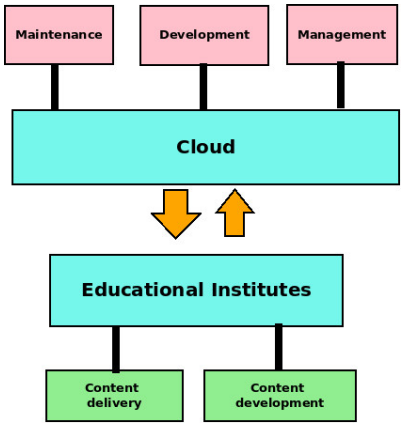


Figure. 10 Separation of roles in e-Learning based on Cloud

6.3 Migrating the e-Learning platform to the Cloud

The proposed solution must ensure the following [4]:

Continuity of information between internal IS and Cloud;

Continuity of information between applications on the Cloud;

Confidentiality of transferred data;

Confidentiality of existing data in IS Campus.

The features needed to integrate cloud-to-cloud and IS-to-cloud applications are the same. Both types rely on technical connectivity, data mapping, different exchange patterns, and so on.

The ESB exchange server and the e-Learning platform management module will need a maximum level of security, which requires to keep them in a private cloud.

The EducDirect application and the two modules related to teachers and students, in the e-Learning application platform, will need more computing power to respond to users requests, which requires to host them in a public cloud to a service provider .

To avoid losing IS Campus data, we have to set up an automatic backup system that records all IS Campus databases and stores them in a private cloud.

To meet our needs, the hybrid deployment model is the ideal solution in this case study [5]. However, this model points out a problem of communication between the private cloud and the public cloud related to security and performance.

In order to exchange fluently the flow of information between the two environments, it is essential to create links between the two clouds (private and public) by setting up a private network to promote

communication between the two systems without passing through the public Internet. This avoids all problems of failure, deceleration and safety.

Our proposed new architecture for e-learning based on cloud computing and the ESB bus is shown in Figure 11:

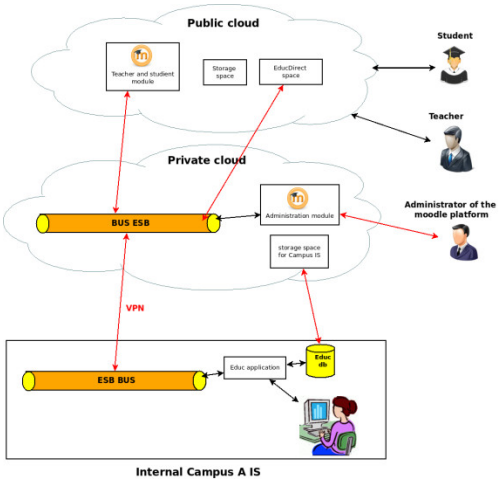


Figure. 11 E-learning model based on cloud computing and ESB

7 E-Learning Cloud Based Architecture

The advantages of the proposed architecture are:

7.1 Powerful computing and storage capacity:

Cloud computing is an infrastructure in which computing and storage power are managed by numerous distributed servers. Throughout cloud computing, users will connect securely to Internet.

The e-Learning architecture based on cloud puts data in a large number of distributed computers to provide high computing power and huge data storage space.

7.2 High data security:

In the Cloud computing model, the software and data are located on remote servers. Service providers manage unified data, allocate resources, balance load, deploy software, monitor security and perform reliable real-time monitoring, ensuring, as far as possible, user data security.

7.3 Less expensive for students and teachers:

E-Learning users do not need high-end configured computers to run e-Learning applications.

They can run applications from the Cloud through their devices (PCs, mobile phones, tablets) having a minimal configuration with the Internet connectivity. Since data is created and accessible in the cloud, the user does not need to spend money for software licenses, software technical support and large memory for storing data in local machines.

7.4 Centralized data storage:

If the client computer crashes, there is almost no data lost because all data are stored in the cloud.

7.5 Virtualization:

Virtualization is the most important feature of this type of architecture. It is very easy to create a clone of a virtual machine so that the downtime of clouds should be greatly reduced.

7.6 High availability:

Due to the integration of massive storage and high-performance computing power, this system can provide a better quality of service. The cloud computing system can automatically detect node failure and exclude it, which does not affect the normal operation of the entire system.

7.7 Easy Implementation and Update:

Updating, can be performed in a short period. Therefore, we will pay more attention to learner's needs, instead of devoting time to implementation and maintenance of the infrastructure.

8 Conclusion

Through this research, we were able to set up a middleware architecture, for the improvement of the governance of an academic institution by using the ESB integration architecture in its various distributed campuses.

The goal is to provide a modern and scalable architecture that combines cloud computing with the ESB application integration architecture to be used in higher education.

We have shown that the benefits of integrating an e-learning system into the cloud can be putted forward for their flexibility and scalability of resources such as high availability, storage, and IT Access to the network. Thus, the cost is calculated according to the number of servers used depending to the number of students.

Finally, the proposed architecture offers interesting possibilities for researchers who wish to deepen and explore the integration of applications of an information system.

By way of perspective, we will improve this architecture by setting up an e-Learning platform at a national level, so to be accessible by the different higher institutions in Morocco.

REFERENCES

[1] Bernard Manouvrier and, Laurent Ménard , “Application Integration: EAI, B2B, BPM and SOA”,ISBN: 978-1-84821-088-2, chapter 5, Published Online: 27 JAN 2010.

[2] BumSuk Jang and Young-guk Ha, “Improvement of Schema-Informed XML Binary Encoding Using Schema Optimization Method”,Computers, Networks, Systems, and Industrial Applications, Springer, 2011.

[3] Chappel, Dave , “Enterprise Service Bus”, O'Reilly,2004.

[4] Christian Baun, Marcel Kunze, Jens Nimis, and Stefan Tai, "Cloud Computing Web-Based Dynamic IT Services," in Cloud Computing.: Springer Berlin Heidelberg, 2011, pp. 15--48.

[5] Dimitrios Zissis and Dimitrios Lekkas, “Addressingcloud computing security issues”, Future GeberationComputer Systems, pp. 583-592, 2012.

[6] Faouzia MESSAOUDI , "DISPOSITIFS DE E-LEARNING : QUELS USAGES POUR AMELIORER LA FORMATION AU MAROC ? Du premier baromètre national du e-Learning à l'étude de cas de la formation continue" , pp.58-95, 2013.

[7] Fenouillet, F & Déro, M, " Le e-learning est-il efficace ? Une analyse de la littérature anglo-saxonne, in Savoirs ", vol. 12, n° 3, pp. 88-101,2006.

[8] FLETCHER J. D, " Effective-ness and cost of interactive videodisc instruction », Machine Mediated Learning, pp. 361-385, 1991.

[9] Guillaume SUPERCZYNSKI, "Alfresco 3.4 Travail collaboratif et GED avec la plate-forme Share", Editions ENI, 2011.

[10] isga.ma, "GESFO System", 2008. [Online]. Available: http://isga.ma/fr/espace-etudiant.

[11] J.A. Hijar Miranda,Daniel Vazquez Sanchez, Dario Emmanuel Vazquez Ceballos,Erika Hernandez Rubio, Amilcar Meneses Viveros, Elena Fabiola Ruiz Ledezma, "Middleware to integrate heterogeneous Learning Management Systems and initial results", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 5, No. 10, 2014.

[12] Marinela Mircea and Anca Ioana Andreescu, "Using Cloud Computin, in Higher Education: A Strategy to Improve Agility in the Current Financial Crisis, IBIMA publishing", Article ID 875547, 2011.

[13] Mell, P., & Grance, T, "The NIST Definition of Cloud Computing", 2012.[Online].Available: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

[14] Sebastian Pätzold, Sabine Rathmayer, Stephan Graf , " Proposal for the design and implementation of a modern system architecture and integration infrastructure in context of e-learning and exchange of relevant data ", pp. 1-8, 2014.

[15] T Mather, S Kumaraswamy, S Latif,Cloud Security and Privacy, O'REILLY, September 2009

[16] Vivek Kale , "Enhancing Enterprise Intelligence: Leveraging ERP, CRM, SCM, PLM, BPM, and BI", ISBN: 978-1-4987-0597-4, CRC Press, 2016.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Automating the Conceptual Modeling of Data Warehouse in Information System ERP Type

Ayoub Elotmani, Pr. Halima Bouden
Laboratoire d'Informatique, Recherche Opérationnelle et Statistique Appliquée
Université Abdelmalek Essaadi.
Tetouan, Maroc.
Elotmaniayoub@gmail.com, bouden.halima@gmail.com

ABSTRACT

An information system type ERP (enterprise resources planning) aims to control the information flow and ensure the integrity of unique repository. However, the strategic level of this organization is looking to note a knowledge for vocation to decide and forecast. This feasible with the introduction of the business intelligence, which allowing storing historical data and organized in storage space: data warehouse (DWH), reduced to data mart (DM) that are an extract of the DWH adopted to a specific use, it is modelled for OLAP (on line analytical processing) analyses.

This article is part of a critical study, about the three Data Warehouse design approaches, in order to make the best approach adaptable to an ERP database. On the order hand, to study the proposals of the automation of the phases of the design of the warehouse.

Keyword: ERP, data warehouse, OLAP, MDA, reducing, Genetic Algorithm, Multiple Corresponding Analysis.

1 Introduction

The information system type ERP collect all operational data of the company in a unique relational database in the logical sense. The need for use of these data in the decision-making process, business intelligence that are dedicated to the management of the company, are based on relational sources to build multidimensional form patterns, easily accessible by decisions-makers, and which adapts well to the OLAP analyses. They are generally group them into two storage spaces: data warehouses (DWH), otherwise reduced to data mart (DM) for a particular use.

Small and medium-sized companies do not allow the high cost of a warehouse. As a result, they build their DM directly on their transactional basis OLTP. The present work is concerned with this alternative. It offers a benchmark of the automatics steps to aid the construction of DM's star schema. To develop the best approach to construction of the patterns of data mart dedicated specifically to the relational source of information system ERP type. In order to make proposals for improvement and evolution.

This article is organized as follows: in section 2, we briefly presenting the Enterprise resource planning ERP, and the data warehouse. In addition, the description of the three approaches of design of the data warehouse. A comparative study of these three approaches is presented in section 3. Section 4, is devoted

DOI: 10.14738/tmlai.54.3330
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3330

to the detailed presentation of the process of developing a multidimensional schema that meets the limits of the comparative study.

2 From Relational to Multidimensional

2.1 ERP

ERP or "Enterprise Resource Planning" software can be defined as an application computer customizable, modular and integrated, which aims to integrate and optimize the process of management of the company by providing a single repository and relying on standard management rules. [15].

In General, an ERP system contains the following general characteristics:

- An integrated system that has the feature to integrate several functions of the company [3]. It extends to the whole of the business processes, the MRP concept developed in the 1970s for industrial management.
- A modular tool composed of a set of application modules (usually signed by the same publisher) that shares a single, logical database. Integrating these modules around a single repository of data is likely to ensure workflow collaboration.
- A "configurable" system allows combining standardization and adaptation to the company [3]. He opposed the proprietary software developed for the specific needs of the company. It allows quick adaptation to the operating rules offering tools of development or customization of application add-ons [8].

Therefore, the ERP is software integrated with the characteristic to cover several or all functions of the enterprise, each managed by a function module, the integration of these modules around one unique repository of data.

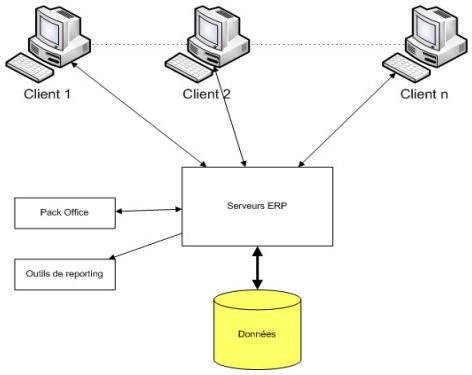


Figure 1: Third Architecture of an Enterprise Resource Planning

2.2 Data WareHouse

The classic definition of a warehouse given:

"A data warehouse is a collection of subject-oriented data, integrated, non-volatile, histories and organized for the support of a decision-making process."

We detail these features:

- Subject-Oriented: data warehouses are organized by topic. For example, a chain of food stores organizes data from its inventory to sales carried out by product and store, during a period.
- Integrated: The data from different sources must be integrated before their storage in the data warehouse. Integration (for example, put in correspondence of formats), to have a consistency of information
- Non-volatile: the difference of the operational data, data warehouse are permanent and cannot be changed. The refreshment of the warehouse is to load new data, without changing or losing existing ones.
- Time-Variant: the evolution of the data is essential for the decision-making process, which, for example, uses prediction techniques based on past trends to predict future developments.

2.3 State of the art and related work

This work is part of the construction of multidimensional data warehouses. The construction of a DWH made from the data of an organization.

The decision-making information system literature proposes three classical approaches to data warehouse design: (i) top-down approach, (ii) bottom-up approach, and (iii) mixed approach.

2.3.1 Top-down approach

The top-down approach is used when economic problems are well known. This approach provides a synergy between business topics and provides a unique version of the truth.

The top-down approach is consistent with Bill Inmon's view that the data warehouse must respond of requirements of all organization's users, not just a certain group.

This type of approach adopted in [7] where the author is based on the identification of the facts and the attributes of the dimensions to elaborate the set of star schemas. Of course, this design method produces a solution that responds well to analytical needs Expressed by future users.

Then it allows constructing multidimensional schemas from the analysis of the needs of the decision makers, the problems of conformity between the OLAP needs (On-line Analytical Processing) and the source of the ERP are manipulated when loading the data in the warehouse.

This method is based on three essential phases:

(i) Data collection, (ii) needs specification, and (iii) formalization of needs

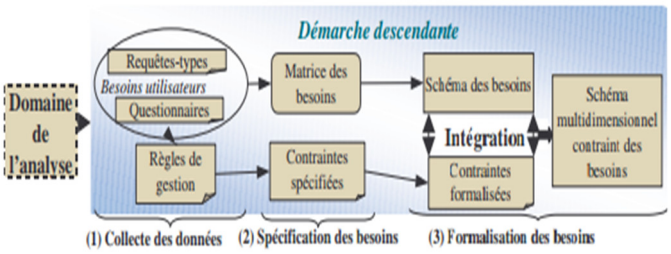


Figure 2: schematic of the top-down approach

Data collection

The data collection phase consists of determining the initial requirements by defining the relevant OLAP queries through questionnaires in order to better position the need.

Requirements specification

This phase allows analysing and organizing data in queries, and translating the easy to integrate in the multidimensional schema constraints management rules.

Formalisation of needs

After collecting the data and specifies the needs of users, the remaining phase is to formalize these multidimensional schema needs. In addition to the definition of the multidimensional facts of this representation must meet the decision-making needs sets.

This construction goes through four steps:

Definition of the facts

Identical measures analysed through parameters collection automatically.

Definition of the dimensions

Each setting characterized by the same measures is automatically grouped into a dimension. Enriched by the addition of attributes low.

Definition of hierarchies

The Organization of the parameters grouped in each dimension in hierarchies.

- Multidimensional schema definition.

According to needs, collection dimensions are automatically assigned to the facts that fit.

2.3.2 The bottom-up approach:

The bottom-up approach is based on the information contained in the data source. It is a flexible method that allows the organization to go further with lower costs, by creating independent datamarts from the data source schema (E / A) relationships. In this case, it is necessary to have a designer of the multidimensional schema with a dual competence computer and trades. For it is the designer of the multidimensional schema that must elaborate the different interests of the organization by analyzing the relational schema.

This approach is consistent with Ralph Kimball's view that the data warehouse should be easily understood by users and provide the correct answers as soon as possible.

The construction of the data warehouse schemes is based on the model of the existing relations between entities this type of approach is adopted by [6] and [9]. This approach considers that the relevant information is in the data source.

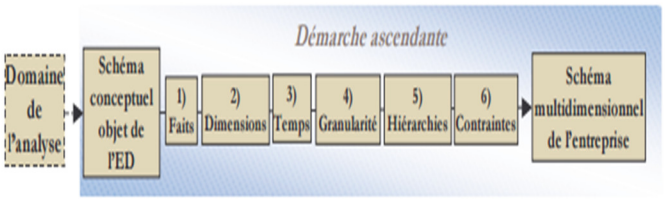


Figure 3: schematic of the bottom-up approach

2.3.3 The Mixed Design Approach:

This type of approach has been adopted in [10] and [2], consists in combining these two design approaches. The source-guided design phase automatically generates candidate DWH schemas. These schemes consist in producing potential facts based on the entities and associations of the source model. This approach has also been taken up by [18], which generate ideal schemes by fusion of analytical needs and then validates these ideal schemes obtained by projection on a power source.

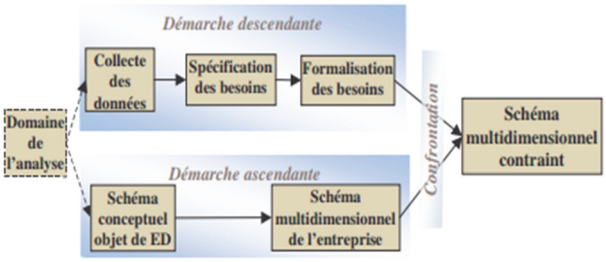


Figure 4: schematic of the mixed approach

Thus, the top-down approach builds candidate schemes that respond well to queries expressed by end users. On the other hand, the bottom-up approach is based on the existing relationships between the data of the relational source and follow a more structured approach to design the company's decision database. On the other hand, the mixed approach is the combination of the two approaches, which consists in constructing multidimensional schemas based on the requests of the decision-makers; on the other hand, it is based on the relational schema source.

The focus of this study is criticism of the top-down approach based on OLAP needs, and the bottom-up, mixed approach to DM design, which is based on the source (ERP data). For drawing the best approach, in order to make improvements.

3 Comparative study of DM design approaches

Reporting and analysing data are the main mission in the process of designing the data warehouse, in fact taking into consideration the source of data is necessary, however the decision makers will be confused while designing the data warehouse so it is such a priority and necessity to show that if neglecting the source . The data will be unstructured in terms of relationships between its characters. Hence, that will lead to the fact that is meaningless for the decision makers to contribute in such cases till the data is structured.

In designing the data warehouse guided by the needs of decision makers, the source data did not taken into consideration. Is right, only the needs of decision makers are necessary to build the data warehouse scheme. So the multidimensional schema design may become incorrect if the existing relationships

between the data are not clearly identified to the designer. On the other, hand unnecessary for decision makers if the problems of the correspondence with the source relational scheme experienced a total divergence.

The authors of [12] propose a design method based on UML modeling. At the conceptual level, the need of decision makers is translated into an UML class diagram format in order to obtain a multidimensional scheme. This approach had drawbacks because user requirements alone are taken into consideration, and the unavailability of the source data influences the implementation of the conceptual schema produced.

On the other hand, the design guided by source exploits the data of the source whatever the level of relevance. Therefore, this case the perimeter of analysis is sometimes wide which requires time and it is difficult to manage the needs of analysts that can be evolved over time or from decision maker to the others. Therefore, designers must always consult the source relational schema that can be uncontrolled.

To remedy this limitation, the authors of [6] propose to the decision-makers to specify the relevant source data. The disadvantage of this proposal is that designers should consult source databases know that they do not always master them.

This approach shares the relational scheme to build facts for independent data mart. However, there may be problems when integrating datamarts into a coherent data warehouse. We can't find the nodes to link datamarts, to construct a coherent constellation schema of a data warehouse

This approach stems from the needs of the company developed by decision-makers, while the top-down approach aims at integration and consistency of data at the enterprise-wide level.

Another proposition in [11], the authors propose a multidimensional schema model constructed from an UML conceptual schema that represents the source. The designer chooses a node of this model to define a fact. All the facts selected are linked with potential dimensions of this fact. However, in a management information system ERP type the relational schemes are wide. Therefore, in our opinion, it is complex for decision makers to represent dimension hierarchies because of the number of nodes generated.

Despite the importance of the fact-finding step from the company's data model, none of these approaches presented a precise method for doing so. While it is always necessary to review the need of end users. In order to remedy this, the mixed approach proposes combining the two approaches [5] and [16]. Indeed, this approach builds the candidate schemas from the data (bottom - up approach) that is to say the determination of these axes of analysis from the relational sources. Beyond this the multidimensional schemas are based on analytical needs (top-down approach). The designer must compare these two types of schemas to obtain a multidimensional result consistent and responding to the needs of decision makers.

[16] Emphasize on automatizing a suitable solution that lies on the Multidimensional Design by Examples (MDBE), which follows a mixed approach. To generate multidimensional schemas, this method takes as input on the one hand the needs of the decision makers expressed in the form of SQL queries and on the other hand the source of relational data. The querying of the sources is ensured by SQL queries and a knowledge of the relational schema of the source. Therefore, building the multidimensional schema requires an expert to formulate SQL queries and query the data sources.

In the shadow of our knowledge, little work involves directly the decision-maker in the process of developing a multidimensional scheme. Note that the article of [11] proposes to the decision-maker to reduce the star schema generated by the system in order to match the scheme to its analysis needs. However, interactions with the decision-maker remain limited.

4 Proposal

In order to respond on the exiting limits approaches,

We consider that the ascending step is relayed separately but in parallel with the downward step, that is to say, the designer can, in the ascending phase, consult the information collected during the downward step. This is proposed in [1] the authors propose an approach (i) mixed, (ii) unified and (iii) automatic for the design of multidimensional DWH. It is a 'mixed' approach to modeling since the schema of the warehouse must be built from the needs of decision-makers and validated against the data sources.

This approach is called 'automatic' because it proposes a set of rules allowing the automatic generation of the logical and physical diagrams from the conceptual schema. In this sense, Model-Driven Architecture (MDA) offers a set of automatic transformation models. In MDA's advocates the development of three types of models: i) the model of the requirements (Computation Independent Model: CIM). (ii) The Analysis and design model (Platform Independent Model: PIM). (iii) The concrete design or code model (Platform Specific Model: PSM)

The MDA's principles must build a model of the requirements (CIM) that automatically turns this in a sequence of models in order to get a physical model adapted to the chosen platform.

In this proposal the first step is not treated, the two following steps allow to build the independent model (PIM) platforms. Which is divided into two different levels of PIM. The first level represents the conceptual model (PIM1) and allows describing the multidimensional structure of the warehouse [13] and specifying the associated ETL treatments. The second level (PIM2) presents different succession of logical models produced automatically from the (PIM1). Finally, the different logical models are translated automatically into a set of platforms of physical implementation models (PSMs).

4.1 Model PIM

In this proposal, the first step is not treated. The next two steps allow building platform independent models (PIMs). That is divided into two different levels of PIM. The first level represents the conceptual model (PIM1) and allows to describe the multidimensional structure of the warehouse [13] and to specify the associated ETL treatment. The second level (PIM2) presents different succession of logical models produced automatically from (PIM1). Finally, the different logical models are translated automatically into a set of physical implantation platform (PSM) models.

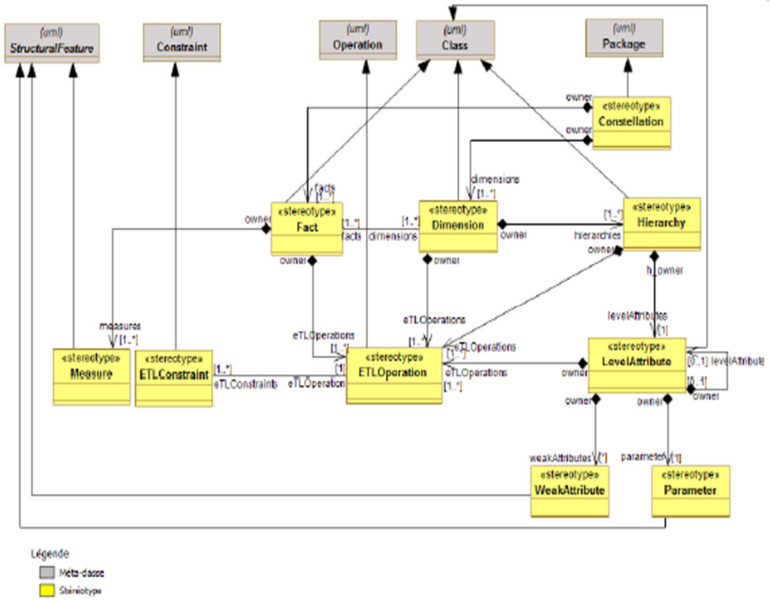


Figure 5 : UML Profile for DWH (DWP)

The UML profile for the DWH explicit in figure 5, composed by all the components of the data warehouse.

In the context of DWHs, static properties are the measurements for the facts, parameters, and weak attributes of the component hierarchies dimensions. The operations concern particularly the warehouse loading processes modelled by the stereotype "ETLOperation" (extension of the meta-class "Operation"). We assume that each fact, dimension, hierarchy, and attribute level must have at least one ETL operation. These operations are described by a list of constraints defined by the stereotype "ETLConstraint". All stereotypes defined in this way have public visibility with the exception of operations and ETL constraints ("ETLOperation" and "ETLConstraint"). The latter must be invisible to the decision-maker. Explain by [1].

The conceptual model (PIM1) is used to represent data that respond to the needs of decision makers in a multidimensional format. This representation is based on the static structure of this model. Therefore, we focus on concepts of fact, dimension, measurement, hierarchy, parameter and weak attribute.

On the other hand, the dynamic aspect of the conceptual model presents the loading operations that concern each multidimensional element. These processes are described using UML processes to validate the multidimensional elements in relation to the different operational sources. After the conceptual model definition (PIM1), the set of rules is defined for the automatic generation of the chosen logical models (PIM2).

4.2 The PSM model

Generally, DWH design stops at logical level. The benefit of this proposal covers the whole cycle of development from formalization of need to physical level. Each constructed logic model is transformed into a logical model through the application of a set of rules of passage to the materialized view.

The physical models (PSM) described by the materialized views are produced by the assembly of the conceptual (PIM1) and logic (PIM2) models. It is a multi-model transformation: one or more input models

are transformed into one or more output model (s). Indeed, each dimension of the conceptual model is translated into an implanted physical dimension. The second input model is the logic model. At this level, each table is transformed into a materialized view.

After the automatic implementation of data warehouse, we always leave a margin where the number of the dimensions are many, for this we thought to look for a method to reducing the dimensions of data warehouse after the conceptual modeling that we adopted as a proposal. According to our studies, we find beforehand a proposal to reducing the multidimensionality of the Olap cubes with the genetic algorithm and the multiple correspondence analysis.

In this context, the authors of [17] propose an approach for reducing the data warehouse's dimensions by identifying a subset of dimensions that may contain the most amount data of the original data warehouse.

This approach uses Genetic Algorithms (GAs) that have proven their efficiency in solving optimization problems, to identify the subset of dimensions to keep. Also used Multiple Correspondence Analysis (MCA) to evaluate different combinations of dimensions proposed by this GAs. It is based on the elimination of certain dimensions of the Olap cubes by merging the groups of candidate facts created in order to be regrouped and replaced by a unique fact.

MCA is a multidimensional analysis method of an array containing qualitative data; it is considered a qualitative data matrix. Where a set of individuals in the lines is described by a set of attributes in the columns. A special data structures for the ACM analysis method including tow table:

- Table of Condensed Coding (TCC): It is a table with n rows and p columns. At the intersection of the row (variable) and the column (individual).
- Complete Disjunctive Table (CDT): This table is built from the TCC and has n rows includes the modalities of variables associated with the m columns.

The MCA offers the possibility of comparing two rows or columns by calculating the distance between them. Such a technique makes it possible to calculate the degree of similarity between these rows or columns. This distance calculation can be done using the distance square to compare two points of individuals (two line profiles) or a column arrangement (two vertical profiles).

In this approach, the goal is to reduce the number of dimensions of a data warehouse. It is a matter of selecting the dimensions contain the most amount information for decision makers from the initial dimensions. In this case, the use of the GAs approach is interesting.

Each individual designates the dimensions of the warehouse, which also represents a possible solution to the problem. Will be represented by a chromosome containing the same number of genes as data warehouse dimensions. The function's adaptation used to evaluate the quality of each individual in the different successive populations corresponds to the calculation of the distances between the individual's terms of the MCA. We can deduce this calculation by the modality of equation to calculate the distance between two terms points. Which represents the degree of resemblance between two individuals, for each modality. After the sum of all the corresponding calculated values of the modalities of all dimensions, we can evaluate the new distance decreases between individuals.

The following algorithm proposed by [17] for this purpose:

```
1. for i from 1 to 10 do
2. for j from 1 to 6 do
3. TCC[i][j] ← rondom(1,0)
4. end for
5. end for
6. for i from 1 to (N=10) do
7. j=1
8. for k from 1 to length(dim1[]) do
9. if (ind[i][j] = dim1[k])
10. then TDC[i][k] ← 1
11. end if
12. end for
13. j[m=1 to 5]=j[m-1] + length (dim[p=2 to 6][])
14. for k[m=1 to 5]=1 tolength (dim[p=2 to 6] []) do
15. if (ind[i][j] = dim[p=2 to 6][k(m=1 to 5)]
16. then TDC[i][k(m=1 to 5)] ← 1
17. end if
18. end for
19. end for
20. for i from 1 to 10 do
21. for j from 1 to 19 do
22. if (TDC[i][j] 1)
23. then TDC[i][j] ← 0
24. end if
25. end for
26. end for
27. for k from 1 to 6 do
28. for j from 1 to 19 do
29. Occ← 0
30. for i from 1 to 10 do
31. som← 0
32. if TDC[i][j] = 1
33. then occ← occ+1
34. end if
35. end for
36. A= occ/N
37. B=(TDC[i][j]-xref)²
38. X=B/A
39. som← som+X
40. end for
41. dist← som/p
42. end for
43. {ind1(i+1), ind2(i+1), ind3(i+1), ind4(i+1)} ← min {ind1(i), ind2(i)…,
ind10(i)}
44. {ind5(i+1), ind6(i+1)} ← crossover ( ind1(i+1), ind2(i+1), pc )
45. {ind7(i+1), ind8(i+1)} ← crossover ( ind3(i+1), ind4(i+1), pc )
46. {ind9(i+1)} ← mutation ( ind1(i+1), pm )
47. {ind10(i+1)} ← mutation ( ind3(i+1), pm )
```

Figure 6: The reduction corresponding algorithm

We can obtain in the example treated by [17] that the reduction of the dimensionality of the OLAP cube using the GA and the MCA is followed by a reduction on the lines of the data warehouse. We also notice the lines that are similar. Become doable to eliminate and replace them with a single line and summarize the measurements of each in the new line. In FIG. X, the algorithm corresponding to the process of identifying the duplicate lines and the replacement by the correspondent.

```
1. while i<=n do
2. for (j,k) from 1 to p' do
3. tab[k] ←CDTm[i][j]
4. end for
5. i++
6. while (j,k)<= p' do
7. if tab[k]==CDTm[i][j] then (j,k)++
8. else i++
9. end if
10. m=1
11. measure=0
12. if k==p' then
13. l[m]=i
14. measure= measure+measure[i]
15. m++
16. end if
```

Figure 7: Corresponding algorithm to the identification of duplicated lines.

5 Discussion

Our study focuses on the automatic construction of multidimensional data warehouse schemes to bypass problems of decision making in ERP-type management information systems. The solution proposed by [1] is adaptable for our case because it responds to the need to design and build a physical data warehouse scheme automatically - this solution is based on the principle of model-driven architecture - for to produce historically, aggregated, business-oriented information, described in a unique and consistent way, mainly based on the basic data of the ERP. Thus, to remedy this problem, the problem of correspondence between schemas built based on user requirements with the relational schema of the ERP, which is sometimes wide. This proposal develops a mixed approach that combines the two top-down and bottom-up approaches.

In the next article, we will discuss a problematic which will improves this work. For processing large amounts of data. These include handling large non-relational databases, data warehouse support. That allows easy access to a Cloud environment. The results obtained can be used to circumvent the problem of decision-making regarding the introduction of ERP in cloud environments.

6 Evolution and Conclusion:

This document presents an ERP as an unavoidable tool that integrates all the company's processes and facilitates the flow of data. The ERP contains many important data for decision-makers. The disadvantage is that an ERP generates only reports of particular problems. For this, we have thought of defining a solution that is based on relational data, in order to build a data warehouse automatically. We take into account the needs defined in the end users.

We have studied the three bottom-up, top-down and mixed-design approaches of the data warehouse, and discuss the boundaries of each. In order to draw the mixed approach as a suitable approach to our case. With the solution presented in [1] whose authors propose a model-driven approach, for the design and automatic construction of a data warehouse.

The future work is to propose a solution for processing large amounts of data. These include manipulating large non-relational databases, data warehouse support. That allows facility to access to a Cloud environment. That goes with the new revolution in ERP implementation in cloud environment.

REFERENCE

[1] Atigui F., Ravat F., Teste O., Zurfluh G., Démarche dirigée par les modèles pour la conception d'entrepôts de données multidimensionnelles. 26éme Journées des Bases de données Avancées BDA, Toulouse, 2010.

[2] Bonifati A., Cattaneo F., Ceri S., Fuggetta A., et Paraboschi S. « Designing Data Marts for Data Warehouse », ACM Transactions on Software Engineering and Methodology, vol. 10(4), p. 452-483. Octobre 2001

[3] Davenport, Thomas H. « Putting the enterprise into the enterprise system ». Harvard business review 76 (4). 1998

[4] Feki J., Hachaichi Y. Du relationnel au multidimensionnel : Conception de magasins de données. Sfax, janvier 2007.

[5] Giorgini P., Rizzi S. et Garzetti M., Goal-oriented requirement analysis for data warehouse design. ACM International Workshop on Data Warehousing and OLAP (DOLAP'05), p. 47-56. Bremen, Novembre 4-5, 2005.

[6] HÜSEMANN B., LECHTENBÖRGER J. et VOSSEN G., Conceptual data warehouse modeling. 2nd international Workshop on Design and Management of Data Warehouse (DMDW'00), p. 6.1 – 6.11, Vol. 28, Suède, June, 2000.

[7] KIMBALL R. The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouses. New York, 374 p, 1996

[8] Lequeux. 1999. Manager avec les ERP. Editions d'Organisation.

[9] Pang C., Taylor, K., Zhang, X. et Cameron, M. Generating Multidimensional Schemata from Relational Aggregation Queries. pp. 584–589, 2004.

[10] Phipps, C., Davis K., Automating data warehouse conceptual schema design and evaluation. 4th International Workshop on Design and Management of Data Warehouse (DMDW'02), p. 23-32, Canada. May 272002.

[11] Pinet. F, et Schneider. M., A Unified Object Constraint Model for Designing and Implementing Multidimensional Systems. Journal on Data Semantics XIII, pp. 37-71, 2009

[12] Prat N., J. Akoka, I. Comyn-Wattiau. "A UML-based data warehouse design method", Decision Support System, Vol.42(3), p. 1449-1473, 2006.

[13] Ravat F. Modèles et outils pour la conception et la manipulation de systèmes d'aide à la décision. Université des Sciences Sociales, Toulouse, décembre 2007.

[14] Ravat F., Teste O., Tournier R., G. Zurfluh. « Querying Multidimensional Data bases ». 11th East-European Conferance on Advances in Databases and Information Systems (ADBIS'07) p.298-313, Varna, septembre 2007.

[15] Reix, Robert, Bernard Fallery, Michel Kalika, et Frantz Rowe. Systèmes d'information et management des organisations. 6e édition. Paris. 2011

[16] Romero. O, et Abelló. A. 2010. Automatic validation of requirements to support multidimensional design. Data & Knowledge Engineering. Pp. 917-942, Septembre 2010.

[17] Semeh B., NAOUALI S., Reducing the multidimensionality of OLAP cubes with Genetic Algorithms and Multiple Correspondence Analysis. The International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015). Pp 452 – 459. 2015

[18] Soussi A., Feki, J., Gargouri. F. Approche semi-automatisée de conception de schémas multidimensionnels valides, Journée Entrepôts de Données et l'Analyse en ligne (EDA'05), Revue des Nouvelles Technologies de l'Information, RNTI-B-1, Cépadues, P71-90. Juin 2005.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

A Novel Compact CPW Multi-StopBand Filter with DGS Integrating Circular Ring Resonator

Elmahjouby Sghir[1], Ahmed Errkik[1], Jamal Zbitou[1], A.Tajmouati[1], Larbi El Abdellaoui[1],
[1]LMEET Laboratory FST of Hassan 1st University, Settat, Morocco
[2]Mohamed Latrach
[2]Microwave group ESEO, Angers, France
elmahjouby.sghir@gmail.com

ABSTRACT

A novel and compact multi-band stop filter with good selectivity and wide upper pass band performance with good rejection based on defected ground structure (DGS) integrating a circular ring resonator to suppress undesired spurious bands frequency or harmonics in microwave and millimeter wave circuits is proposed.

The presented filter has a compact size because the circular ring resonator is integrated in DGS, the both of them are etching in the ground plane.

Index Terms— Defected ground structure (DGS), coplanar line, stopband filter, multi-stop stopband.

1 Introduction

The microwave circuits such as filters design is one of many application domain of DGS structure, which should give excellent performance in terms of response as bandwidth, quality factor and the parasite rejection, also the level of integration as the size, the development and integration of other components and their interconnection, [1-2].

The concept of the defected ground structure (DGS) has been derived from the idea of photonic bandgap structure (PBGs).[3-4].

However, the DGS structure is applied to design many microwave circuits type, such as microstrip antenna [5], dividers [6], couplers, amplifiers, oscillators, [7]. and microwave filters. Although, The most reported structures of filters on DGSs are designed for single-band (Pass or Stop) or used DGS for improving applications [8-9-10], or integrating a periodic cells for extend wide stopbands, [11].

Recently, there are some works and papers about dual-band bandstop filter presented DGS technique, [12–13]. Now, the multi stopband filter is one of the important microwave and millimeter components are highly applied in modern microwave systems because of their ability to suppress no desired bands signals and harmonics, [14-15]. For that a coplanar wave guide structure with DGS technique, and DGS integrating circular ring resonator, are a good condidate to build these kinds of filters with compact size and high performance, to answer the demands of modern communication systems.

DOI: 10.14738/tmlai.54.3331
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3331

In this paper, we have developed a new design concept of multi-band stopband filter BSF using circular defected ground structure (DGS) with shaped coplanar line is investigated for compact single stopband filter (SBF) structure with the response of single resonant element in 20 GHz exhibits the stopband function, and the circular DGS integrating a circular ring resonator (CRR) for tri-band stopbands with compact size for practical microwave and millimeter applications operating between 1 GHz to 100 GHz frequency.

2 Single Band of Initial Dgs Stopband Filter Design

A Stopband filter with DGS circular structure is designed, Figure 1 depicts the schematic of the proposed structure. It consists of a **50Ω** conventional coplanar transmission line, with a signal line width of W=108µm and the gap width, G=60µm.The substrate is with dielectric permittivity of er=11.9 and thickness of H=200 µm, [16].

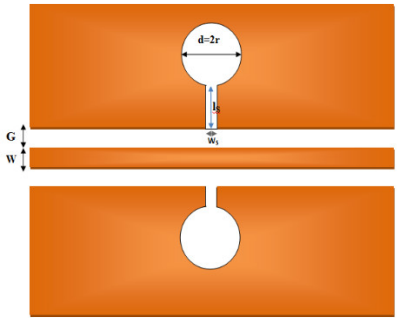


Figure. 1. CPW line with Circular DGS structure

The DGS configuration in this structure is a circular slot etched in the ground plane with a diameter d. This circular DGS is connected to the line gap by a rectangle transverse slot with length of $\mathbf{l_s}$ and width of $\mathbf{w_s}$. The Momentum (an EM solver integrated in ADS Agilent) is used for deriving and studying the filter's electrical performances.

3 Responce of Initial Circular DGS Filter

It is widely known that the DGS circular slot etched plays the role of resonator element and its frequency resonance depend on circular geometry parameters. Figure 2 depicts the response of the CPW DGS Filter.

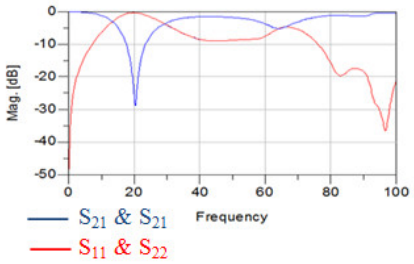


Figure. 2 Simulated S-parameters for the CPW DGS Circular structure

As noted from the S-parameter studies, we have a stop band behavior of the circuit with a good bandstop around 20 GHz.

Moreover, we analyzed the dependence of the rectangle slot on the S-parameters response, the ray of circular etched is kept constant at r = 600µm, also the width of the rectangle slot is fixed at w_s = 60µm.

The length $\mathbf{l_s}$ was optimized for two different value of l_s = 400µm and 500µm. The simulation results obtained with l_s = 500µm are shown in Figure. 3.

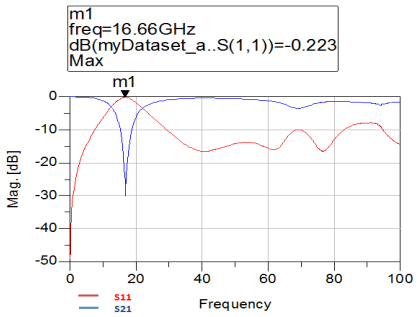


Figure. 3. Simulated S-parameters for the CPW

As a result, there is a significant difference in S-parameter between the initial response structure and this later with the new length ls. We see that the problem is resolved and insertion loss (S21) and return loss (S11) are affecting in the passband, S_{21} is better than −10dB and S_{11} is better than −3 dB in stopband with the new configuration.

4 Modelling of the Circular DGS Stopband Filter

Circular DGS combined with coplanar line causes a resonant response of the structure transmission, and the resonant frequency dependants by circular etched parameters and size of the slot. Its RLC-equivalent circuit of the cell is proposed, and the different parameters can be extracted as follows [17]:

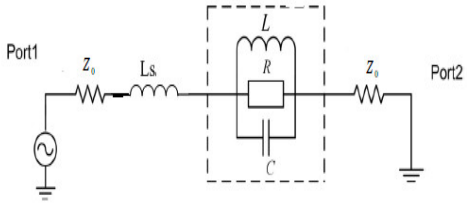


Figure. 4. RLC-equivalent circuit for CPW DGS Filter

$$C = \frac{w_c}{2Z_0(w_0^2 - w_c^2)} \quad (1)$$

$$L = \frac{1}{4(\pi f_0)^2 C} \quad (2)$$

$$R = \frac{2Z_0}{\sqrt{\frac{1}{|S_{11}(w_0)|^2} - (2Z_0\left(w_0 C - \frac{1}{w_0 L}\right))^2 - 1}} \quad (3)$$

Where, Z_0 is the characteristic impedance of the coplanar wave guide line, w_0 is the angular resonance frequency, w_c is the -3dB cutoff angular frequency, and $S_{11}(w)$ is the input reflection coefficient of the equivalent circuit, which can be obtained from EM simulation results.

5 Design of Multi-Band Stopband Filter

Based on the above design and analysis of circular DGS-SBF, tri-bands BSF response was designed on the same first structure.

The initial DGS Stopband Filter design is modified by introducing a circular ring resonator (CRR) with O-configuration to create the tri-bands stopband and to generate the second and third stopband performance, thereby maintaining the compact seize and easy fabrication of structure.

The parameters of the DGS StopBand filter are kept unchanged while dimensions of the circular ring resonator (CRR) are introduced, as shown in Fig. 5.

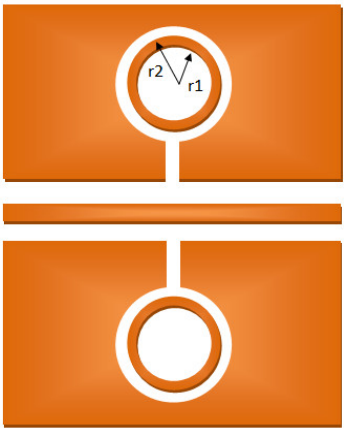


Figure. 5. The proposed device of tri-band stopband filter with O-CRR

The dimensions of the O-configuration of the circular ring resonator (CRR) are as first radium (r_1=409µm), and the second radium (r_2=470µm).

Figure 6 shows the simulated S-parameters of the proposed device of tri-bands stopband filter with O-stub in the circular DGS.

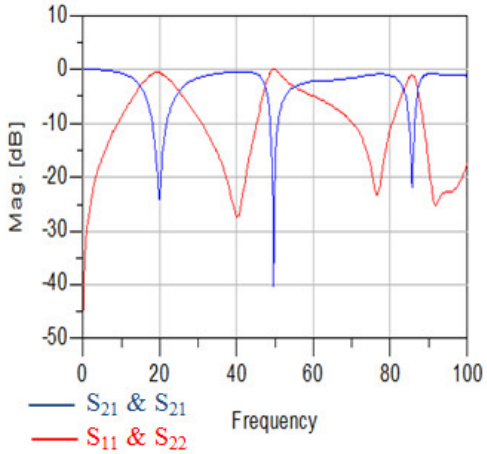


Figure. 6. Simulated S-parameter of the proposed device with O-CRR

As a result, the first stopband resonance is always around of f_{01}=20 GHz, it is approximately centered between 16 GHz and 23 GHz. The three resonance frequencies are measured at f_{01}=20 GHz and f_{02}=50 GHz (f_{02}=2.5 × f_{01}) and f_{03}=86 GHz (f_{03}=2.5 × f_{02}), with respective rejection levels of 25 dB, 40 dB and 22.30 dB.

Similarly, their three return loss S_{11} is better than −3 dB in stopband and offering a drastically improved response performance of stopband filter. So the O-Stub inserted in the DGS of a CPW structure can modify the insertion loss (S21) and return loss (S11) behavior, to get a multiple band stopband after the first one of the circular DGS. Also, the tuning resonance of the second and the third stopband can be achieved by varying the radius of the O-stub in the circular DGS.

6 Conclusion

In this paper a new compact and simple structure using one circular DGS integrating a circular ring resonator with O-configuration have been successfully simulated and applied to the design of compact multi-bands stopband filter in coplanar wave guide technology for many millimeter wave applications, with excellent suppressed, sharp rejection slope and wide stopband rejection. The O-CRR radiuses play a key role in adjusting and tuning the resonance of the second and third stopband center frequency for another different frequencies.

ACKNOWLEDGMENT

We thank Mr. Mhamed Latrach from ESEO Angers in France, for allowing us to use and to perform simulations by using electromagnetic solvers and measurement instruments.

REFERENCES

[1] Jia-Sheng Hong, and Bindu M Karyamapudi, " A General circuit model for defected ground structure in planar transmission lines," IEEE Microwave Wireless Comp. Lett., Vol.10, No.10, pp. 706-708, Oct. 2005.

[2] H A. T. Mendez, H. J. Aguilar, and R.F. Leal, et. al, "Improving frequency response of Microstrip filters using defected ground and defected microstrip structures," Progress in Electromagnetic Research C, Vol.13, pp. 77-90,

[3] 2010.L. H. Weng, Y. C. Guo, X. W. Shi, et al., "An overview on defencted ground structure," Progress in Electromag- netics Research B, Vol. 7, pp. 173-189, 2008.

[4] V. Radisic, Y. Qian, R. Coccioli and T. Itoh, "A novel 2-D photonic bandgap Structure for microstrip lines," IEEE Trans. Microwave and Guided Wave Lett., vol. 8, no. 2, pp. 69–71, Feb. 1998.

[5] J. P. Thakur and P. Jun-Seok, "A new design approach for circular polarize antenna with DGS under the unbalanced feed-lines," 36th European Microwave Conference, pp. 1483-1485, 2006.

[6] J.-J. Koo, S.-M. Oh, M.-S. Hwang, *et al*., "A new DGS unequal power divider," European Microwave Conference, pp. 556-559, 2007.

[7] A.A. Alsanousi, A. Ashoka, W.Jeberson**.** "A Compact Microstrip Low Pass Filter Based on DGS with Shaped Microstrip Line" International Journal of Engineering and Technology Volume 3 No. 6, June, 2013.

[8] Al Sharkawy, M. H., D. Abd El Aziz, and E. Galal, "A miniaturized lowpass/bandpass filter using double arrow head defected ground structure with centered etched ellipse,".

[9] Rehman, S. U., A. F. Sheta, and M. Alkanhal, "Compact bandstop filter using defected ground structure (DGS)," Saudi International Electronics, Communications and Photonics Conference, SIECPC, 2011.

[10] Woo, D.-J., T.-K. Lee, J.-W. Lee, et al, "Novel U-Slot and V-Slot DGSs for bandstop filter with improved Q factor," IEEE Trans. on Microwave Theory and Techniques, Vol. 54, No. 6, 2840–2847, 2006.

[11] Xiao, J.-K., Y.-F. Zhu, and S. F. Jeffrey "DGS lowpass filters extend wide stopbands," Microwaves & RF, Vol. 50, No. 3, 66– 76, 2011.

[12] Jun Wang1, Huansheng Ning1, Qingxu Xiong1, "A novel miniaturized dual-band bandstop filter using dual-plane defected structuree" Progress In Electromagnetics Research, Vol. 134, 397{417, 2013.

[13] Che, W. Q., W. J. Feng, and K. Deng, "Microstrip dualband bandstop filter of defected ground structure and stepped impedance resonators," International Journal of Electronics, Vol. 97, No. 11, 1351–1359, 2010.

[14] Y. Horii and M. Tsutsumi, "Harmonic control by photonic bandgap on microstrip patch antenna," IEEE Microw. Guided Wave Lett., vol. 9, no. 1, pp. 13–15, Jan. 1999.

[15] H. Liu, Z. Li, X. Sun, and J. Mao, "Harmonic suppression with photonic bandgap and defected ground structure for a microstrip patch antenna," IEEE Microw. Compon. Lett., vol. 15, no. 2, pp. 55–56, Feb. 2005.

[16] M. F. Karim, A.-Q. Liu, A. Alphones, X.J. Zhang and A.B. Yu. "CPW band-stop filter using unloaded and loaded EBG structures" IEE Proc.-Microw. Antennas Propag., Vol. 152, No. 6, December 2005.

[17] Ahn,D.,J. S. Park,C. S. Kim,J. Kim,Y. X. Qian,and T. Itoh, "A design of the low-pass filter using the novel microstrip defected ground structure," IEEE Trans. Microw. Theory Tech., Vol. 49, No. 1,86–93,2001. Progress In Electromagnetics Research Letters, Vol. 24, 99–107, 2011

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Intelligent System for the Management of Resources Dedicated to Humanity

[1]S. Chouya, [2]S. Bourekkadi, [3]S. Khoulji, [4]K. Slimani, [5]M.L. Kerkeb,
[1,2,3,5]Laboratory Information system engineering,
ABDELMALEK ESSAADI University, National School of Applied Sciences, Tetouan, Morocco.
[4]Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco
chouya.saad@gmail.com, sisalmane@gmail.com, khouljisamira@gmail.com; slimani.uit@gmail.com, kerkenml@gmail.com, acmlis.conference@gmail.com

ABSTRACT

Our work consists, one way or another, in projecting the light on the intensive need for the reasonable management of water resources. According to the latest studies and statistics, Morocco will soon face a serious crisis if it has not taken the necessary precautions and decisions to deal with the dangers linked to water resources.

The solution we have developed is not costly at the material / financial level, since it does not imply to the client to have a smartphone or an internet connection, but precisely the possession of the client or other members Of his family of a mobile phone of whatever nature. From this we can assume that our solution adopts the communication network technology of mobile phones without imposing exclusivity on a company without others and that will be totally free.

Among the positive results that our work will bring is the creation of an atmosphere of relaxation and satisfaction among the citizens, through continuous communication and interaction with them, to make them more satisfied with the quality of services. Moreover, this solution is one of the policies that call for the preservation of water resources and the restoration of citizens' confidence in drinking water companies, and on the other hand we will contribute to the positif use of Technology and modern means of communication, and the humanization of our information systems.

To conclude, we believe that these ideas will have a favorable influence on all poles, whether citizens, countries or future generations.

Keywords: Water, Decision-making system, Water ressources, Humanization.

1 Introduction

Morocco, like virtually the entire world, suffers from a threat to its stability and viability. This problem manifests itself in a remarkable decrease in water resources and in its poor management, which puts us face to face with the Problem: how can we fight against the waste of water, and the irresponsible management of this vital element?

DOI: 10.14738/tmlai.54.3332
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3332

Current research and studies on water resources indicate that the dangers associated with this issue need to be targeted and that this threat needs to be unified in order to address this threat through the use of all possible means can help streamline water use.

Morocco has committed itself on a strategy aimed at preserving water resources, among the methods that the state has to adopt in this respect, is the policy of construction of the dams which was considered among the political wise ones, the sensitization of the Citizens to better water management with written, visual and audiovisual supports.

As researchers in the field of computing and development, we must engage in this noble cause that seeks to ensure a decent existence and fair and equitable sharing of this "water" wealth.

Among the barriers to our work is the insufficient data on the rate of water consumption per house, since the water company concerned refuses to collaborate or answer our questions. On the other hand, some citizens provided us all the data and information on their monthly water consumption. In addition, scientific research in this area was unavailable and insufficient or inaccessible, which forces us to multiply our efforts to make it happen.

The implementation of our solution is not costly and does not require much effort from water supply companies or citizens. While our solution will pay dividends and benefits to all participants, it will prevent it from falling into these accidents and clashes between customers and water supply companies, taking what's going in Tangier as an example.

2 Information systems are needed

Man has become possessed and enslaved to computer systems, since these systems help man to perform several tasks and professional or personal activities, man uses this effective tool to perform the least tasks.

Now, the time has come to engage these information systems to control our water consumption and correct some unacceptable behavior in order to move us towards the right path, one that will guarantee the next generations a decent existence and will save the Money that can be used in other things and in other fields, instead of spending them to pay for costly water consumption bills, which causes the citizens to suffer.

The above clearly shows that the information systems if well used, they become an indispensable asset to promote security and stability and serve humanity.

3 The need for governance of water resources management

- Good vision
- Specification of obligations
- Specification of duties
- Responsibility of organizer /citizen
- Harmonization and coherence to establish public interest
- The right intervention

We are convinced that all of the above aims to create a collective consciousness that affects citizens a priori, on the need to invest our efforts and our skills to preserve water and the achievement of objectives while minimizing costs as much as possible.

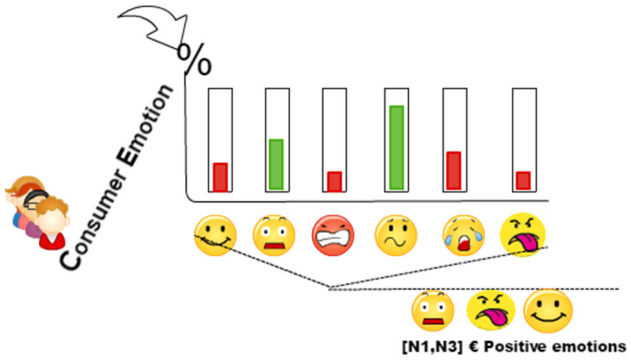


Figure 21 Consumer Emotion

4 Proposition and results

An intelligent information system seems to be an adequate and effective solution that will bring its benefits, since half of the citizens we have questioned have confirmed that such a solution will be a positive step towards creating an atmosphere of satisfaction And trust since they will be informed of the details of the rate of their consumption before the arrival of the bill, which will limit any kind of complication and complaint.

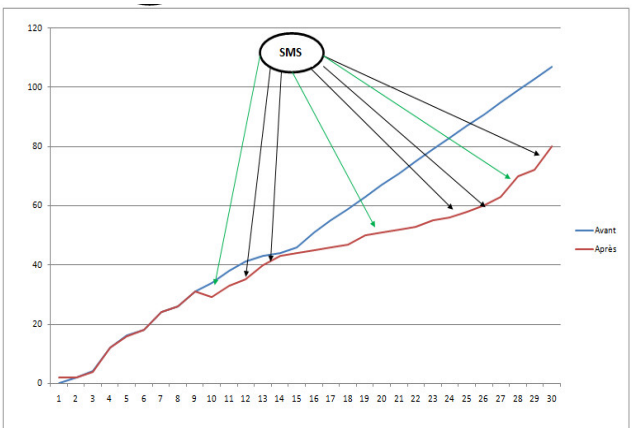


Figure 22 Consumption before and after using our solution

5 Design and architecture of proposed information system

At the level of the user's diagram design, the system has four factors, the information system manager, who decides the customer's need through the intermediary call center, this second factor makes the customer relationship management, The third factor is the technician, that is charged for the development and the improvement of the system, finally the customer the fourth factor, we can even consider it the core of the application. See the figure below which explains in brief the tasks that each factor can do, also the links between these factors.

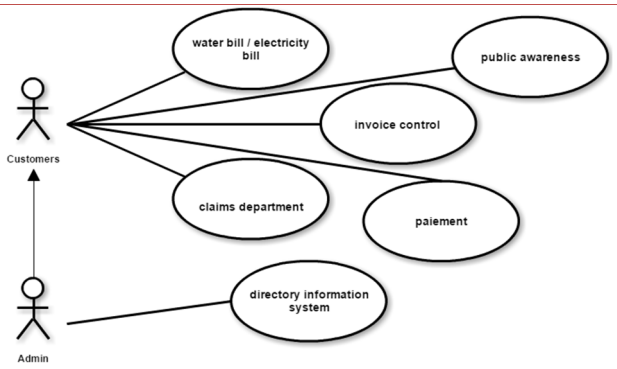


Figure 23 Use Case Diagram

The system will allow its user to perform several tasks, such as consultation of water and electricity bills, online payment of these bills, communication with the call center of the Drinking water and electricity distribution company, file applications and complaints, follow up on individual consumption, as it can also enjoy the e-documents "electronic documents" space without being forced to move to the agency.

Also note that the system is able to sensitize and train customers on the management of water resources and electricity, an adaptive awareness with each type of customers (home consumption, Hotel, company ...) see the figur below that presents a simple architecture of the presented system.

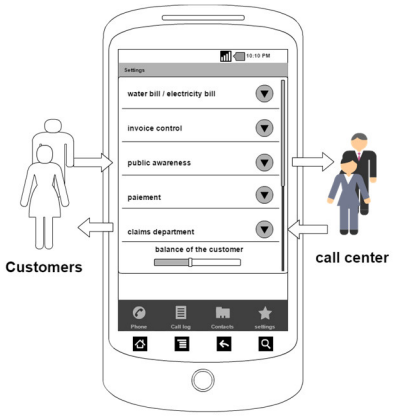


Figure 24 Common Platform of Communication

6 Discussion

We propose a set of algorithmic tools to guarantee – before its deployment - that an application, once installed on a machine, will be able to have a level of humanized communication with its user avoiding the maximum of any source of client dissatisfaction.

Today, it is recognized that new information and communication technologies (ICTs) play an important role in supporting and ensuring the majority of activities carried out within a company, particularly in the field The management of the resources attached to man, which is the object of the continuity of this modest work.

To meet these needs, we rely on the concepts of the man-machine relationship; Our next task is to provide an applicable and realistic solution in the northern region of Morocco.

7 Conclusion

Water is the symbol of life and fertility, it is indispensable, No living being on planet Earth can survive without it. It is used by both humans and ecosystems throughout their cycles. Water is the main constituent of our body, A human being is composed of more than two thirds of water, and a loss of more than 15% of our water can cause dehydration and death. Starting from this observation, water is even sought on Mars, and through its explorations Man seeks to know if he is alone in the universe. This may not be the case, but what remains certain is that man is alone in the face of his responsibilities.

ACKNOWLEDGEMENTS

We would like to express our appreciation to Prof. Dr S.Khoulji and M.L.KERKEB for their great encouragement, moral supports and their guidance, without their valuable assistance, this work would not have been completed.

The authors would also like to address special thanks to JOURANI MOHAMED and ALI BOUAZZAZ for their assistance with particular technique, and for their comments that greatly improved the manuscript.

This research was supported by the grant of National Center for Scientific and Technical Research (CNRST-Morocco): No. 6UAE2015 and 757UIT for the first and the third authors.

REFERENCES

[1] E.B. Foa & E. Meadows, « Psychosocial treatments for posttraumatic stress disorder: A critical review », Annual *Review of Psychology*, 48: 449-480, 1997.

[2] S.Bourekkadi, S.Khoulji, M.L.Kerkeb ,A.Mabrouk "psychologie informatique et son impact sur le comportement humain" *International Journal of Innovation and Applied Studies*, vol. 14 No. 2 Jan. 2016, pp. 543- 548.

[3] KAMATE, M. Z. Etude des troubles psychiatriques chez les emigres dans le service de psychiatrie du chu point «G». 2009.

[4] Sebastiani, F. Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47. 2002.

[5] Howell, D., Bestgen, Y., Yzerbyt, V., & Rogier, M. Méthodes statistiques en sciences humaines. 2008.

[6] Stafford, J., & Bodson, P. L'analyse multivariée avec SPSS. Puq. 2006.

[7] K.Slimani,S.Bourekkadi, R.Messoussi, S.Khoulji, N.Bezad, M.L.Kerkeb "Design of an Awareness System on Sexual and Reproductive Health Based on the Recognition of Emotional Facial Expressions", *International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC)*, March 16 Volume 4 Issue 3 , ISSN: 2321-8169, PP: 193 – 196.

[8] Skinner, E. A., & Belmont, M. J. Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology*, 85(4), 571. 1993.

[9] Day M, Semrad EV. Schizophrenic reactions. In : Nicholi Jr AM (Ed.) The Harvard Guide to Modern Psychiatry. Cambridge, Mass: Belknap Press of Harvard University Press; 1978. p. 199–241.

[10] Goodenow, C. The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools*, 30(1), 79-90. 1993.

[11] Kandel, E. R. Un nouveau cadre conceptuel de travail pour la psychiatrie. *L'Evolution psychiatrique*, 67(1), 12-39. 2002.

[12] Allais, M. La psychologie de l'homme rationnel devant le risque: la théorie et l'expérience. *Journal de la Société de Statistique de Paris*, 94, 47-73. 1953.

[13] Michel Volle, Analyse des données, Economica, 4e édition, 1997 (ISBN 2717832122)

[14] Kahana RJ. Psychotherapy: models of the essential skill. In : Bibring GL (Ed.) The Teaching of Dynamic Psychiatry : A Reappraisal of the Goals and Techniques in the Teaching of Psychoanalytic Psychiatry. Madison, Conn: International Universities Press; 1968. p. 87–103.

[15] De Castro, L. N., & Timmis, J. Artificial immune systems: a new computational intelligence approach. Springer Science & Business Media. 2002.

[16] Schwartz JH, Kandel ER. Modulation of synaptic transmission: second-messenger systems. In: Kandel ER, Schwartz JH, Jessell TM (Eds.) Essentials of Neural Science and Behavior. East Norwalk, Conn: Appleton & Lange; 1995. p. 243–67.

[17] Giorgio Conti et Carlo Perelli, “Governing tourism monoculture: Mediterranean mass tourism destination and governance networks”, dans Peter M. Burns et Marina Novelli (dir.), *Tourism and Politics. Global Framework and Local Realities, Elsevier*, 2007.

[18] Turban, E., Aronson, J., & Liang, T. P. Decision Support Systems and Intelligent Systems 7 “” Edition (pp. 10-15). Pearson Prentice Hall. 2005.

[19] Roy, N., & Garon, R. Hors thème Étude comparative des logiciels d’aide à l’analyse de données qualitatives: de l’approche automatique à l’approche manuelle. Recherches qualitatives, 154. 2013.

[20] Wanlin, P. L’analyse de contenu comme méthode d’analyse qualitative d’entretiens: une comparaison entre les traitements manuels et l’utilisation de logiciels. *Recherches qualitatives*, 3, 243-272. 2007.

[21] Benzécri, J. P. L'analyse des données (Vol. 2, p. l). Paris: Dunod. 1973.

[22] Tiberghien, G., & Beauvois, J. L. Domination et impérialisme en psychologie. *Psychologie française*, 53(2), 135-155. 2008.

[23] Bergadaà, M. Le temps et le comportement de l'individu Deuxième partie. *Recherche et Applications en Marketing*, 4(1), 37-55. 1989.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Improving Small File Management in Hadoop

[1]O. Achandair,[2]M. Elmahouti, [3]S. Khoulji, [4]M.L. Kerkeb,
[1,2,3,4] *Information System Engineering Resarch Group*
Abdelmalek Essaadi University, National School of Applied Sciences, Tetouan, Morocco.
o.achandair@gmail.com, mouradelmahouti@gmail.com,khouljisamira@gmail.com,
kerkebml@gmail.com, acmlis.conference@gmail.com

ABSTRACT

Hadoop, considered nowadays as the de-facto platform for managing big data, has revolutionized the way customers manage their data. As an open-source implementation of map-reduce, it was designed to offer a high scalability and availability across clusters of thousands of machines. Through it two principals' components, which is HDFS for a distributed storage and MapReduce as the distributed processing engine, companies and research studies are taking a big benefit from its capabilities.

However, Hadoop was designed to handle large size files, so when it comes to a large number of small files, the performance can be heavily degraded. The small file problem has been well defined by researchers and Hadoop community, but most of the proposed approaches only deal with the pressure caused on the NameNode memory. Certainly, grouping small files in different possible formats, that are most of time supported by the actual Hadoop distribution, reduce the metadata entries and solve the memory limitation, but that remain only a part of the equation. Actually, the real impact that organizations need to solve when dealing with lot of small files, is the cluster performance when those files are processed in Hadoop clusters. In this paper, we proposed a new strategy to use efficiently some one of the common solution that group files in a MapFile format.

The core idea, is to organize small files files based on specific attributes in MapFile output files, and use prefetching and caching mechanisms during read access. This would lead to less calls of metadata from the NameNode, and better I/O performance during MapReduce jobs. The experimental results show that this approach can help to obtain better access time when the cluster contain massive number of small files.

Keywords - Cloud Hadoop, HDFS, Small Files, SequenceFile, MapFile

1 Introduction

Apache Hadoop, an open-source framework initiated by Yahoo!, developed for storing and processing data at a large scale on clusters of thousands of commodity hardware. Meanwhile, it offers a reliable, scalable, and fault tolerant platform deployed by many big companies such as Facebook, Google, IBM, Twitter and others to manage Terabytes to Petabytes of data.

Hadoop was designed to handle large files, especially when traditional systems are facing limitations to analyze this new data dimension caused by the present data explosion. However, large files are not the

DOI: 10.14738/tmlai.54.3333
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3333

only kind of files that Hadoop deployments needs to manage, the diversity of all kind of data becomes a standard in most of big data platform. There are many fields that produce tremendous numbers of small files continuously such as analysis for multimedia data mining [6], astronomy [7], meteorology [8], signal recognition [9], climatology [10,11], energy, and E-learning [12] where numbers of small files are in the ranges of millions to billions. For instance, Facebook has stored more than 260billion images [13]. In biology, the human genome generates up to 30 million files averaging 190KB [14].

1.1 HDFS – Storage Layer

HDFS, The Hadoop distributed file system provides high reliability, scalability and fault tolerance. It's designed to be deployed on big clusters of commodity hardware. It's based on a master-slave architecture, the NameNode as a master and the DataNodes as slaves. The NameNode is responsible for managing the file system namespace, it keeps tracks of files during creation, deletion, replication [3] and manages all the related metadata [4] in the server memory. The NameNode splits files into blocks and sends the writes requests to be performed locally by DataNodes. To ensure a fault-tolerance system, blocks replicas are pipelined across a list of DataNodes. This architecture as shown in "Fig.1", with only one single NameNode simplifies the HDFS model, but it can cause memory overhead and reduces file access efficiency when dealing with a high rate of small files.

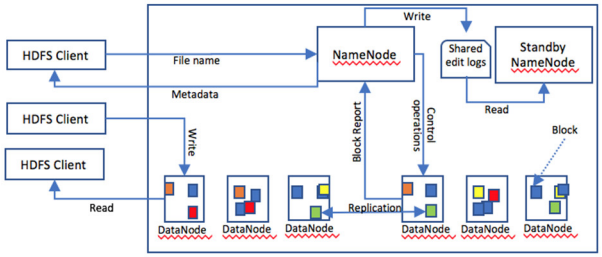


Figure. 1: HDFS Architecture

1.2 MapReduce – Processing Engine

In the current version of Hadoop, Google re-architected the processing engine to be more suitable for most of big data applications needs. The major improvement of Hadoop was the introduction of a resource management module, called YARN, independently of the processing layer. This brought significant performance improvements, offered the ability to support additional processing models, and provided a more flexible execution engine. Because of its independency architecture, existing MapReduce applications can run on YARN infrastructure without any changes.

The MapReduce program execution on YARN can be described as follows:

i. A user defines an application by submitting its configuration to the application manager
ii. The resource manager allocates a container for the application manager
iii. Resource manager submits the request to the concerned node manager
iv. The Node manager launches the application manager container
v. The application manager gets updated continuously by the node manager nodes, it monitors the progress of tasks

When all the tasks are done, the application manager unregisters from the resource manager, like so, the container can be allocated again, See "Fig. 2".

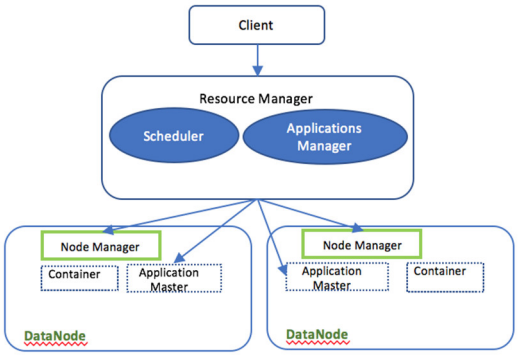


Figure. 2: YARN – Yet Another Resource Manager

The rest of this paper is divided into the following, section 2 lists the existing solutions in the related literature. Section 3 present the proposed approach. Section 4 is allocated for our experimental works and results. Finally, Section 5 for conclusion and expectation.

2 Related Work

To deal with the small file problem, numerous researchers have proposed different approaches. Some of those efforts have been adopted by Hadoop and are available for use natively, more precisely, Hadoop Archives (HAR Files) and SequenceFile.

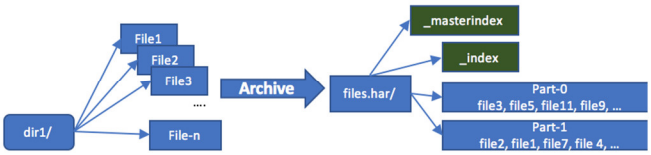


Figure. 3: HAR File Layout

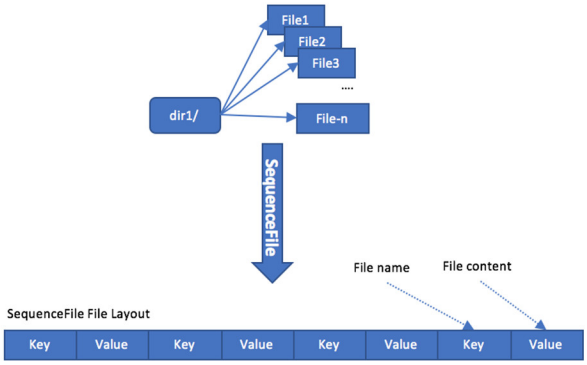


Figure 4: SequenceFile File Layout

Hadoop Archive packs small files into a large file, so that we can access original files transparently, see "Fig. 3". This technique allows more storage efficiency, as only metadata of the archive is recorded in the namespaceof the NameNode, but it doesn't resolve other constraints in terms of reading performance.Also, the archive cannot be appended while adding more small files. The SequenceFile technique is to merge a group of small files in a flat file, as key-value pairs, while key is the related file metadata and value is the related content, see "Fig. 4".

Unlike the HAR files, the SequenceFile supports compression, and they are more suitable for MapReduce tasks as they are splittable [15], so mappers can operate on chunks independently.However, converting into a SequenceFile can be a time-consuming task, and it has a poor performance during random read access.

To improve the metadata management, G. Mackey et al. [16] merged small files into a single larger file, using the HAR through a MapReduce task. The small files are referenced with an added index layer (Master index, Index) delivered with the archive, to retain the separation and to keep the original structure of files.

C. Vorapongkitipunet al. [17] proposed an improved approach of the HAR technique, by introducing a single index instead of the two-level indexes. Their new indexing mechanism aims to improve the metadata management as well as the performance during file access without changing the implemented HDFS architecture.

Patel A et al. [18] proposed to combine files using the SequenceFile method. Their approach reduces memory consumption on the NameNode, but it didn't show how much the read and write performances are impacted.

Y. Zhang et al. [19] proposed merging related small files according to WebGIS application, which improved the storage effeciency and HDFS metadata management, however the results are limited by the scene.

D. Dev et al. [20] proposed a modification of the existing HAR. They used a hashing concept based on the sha256 as a key. This can improve the reliability and the scalability of the metadata management, also the reading access time is greatly reduced, but it takes more time to create the NHAR archives compared to the HAR mechanism.

P. Gohil et al. [21] proposed a scheme for combining small file, merging, prefetchingthe related small files which improves the storage and access efficiency of small files, but does not give an appropriate solution for independent small files.

3 The Proposed Approach for Small File Management

The core idea behind our approach is to combine different clients' files that contain sets of small files, when relevant, and merge them through a merger process, to be stored in an optimal way before closing the current SFA connection. This implementation has been achieved in our previous work, See "Fig 5", as the main task of the Small File Analyzer Server. In the current research, we improved the proposed SFA server, to handle other modules that offered us the possibility to consider other parameters during the merger process.

We introduced a sorter process, that can operate as an independent module in the SFA server. The compressor module is another added module, that allowed us to compress merged files that are no longer used, or rarely called, such a way we can benefit from a considerable benefit of storage capacity. Also, we used a prefetching and caching technique to boost the performance while reading the same small files.

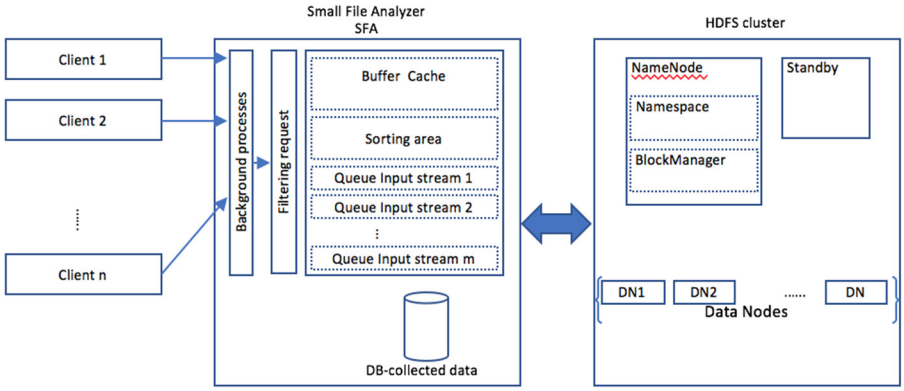


Figure. 5: SFA Architecture

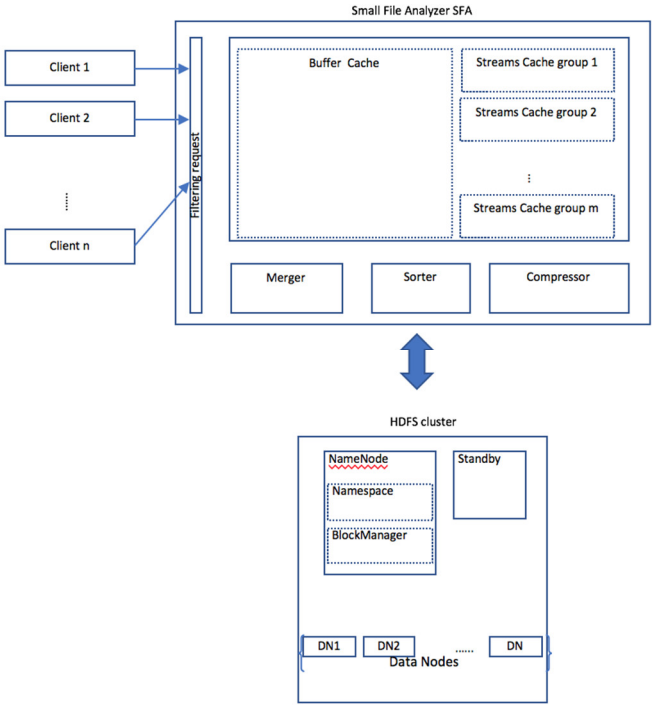


Figure. 6: SFA improved Architecture

For simplicity, the clients' files are stored in the archaically files system of the SFA server, connecting clients' streams directly to the merger is not covered in the scope of the current study. However, its considered to maintain the same concept when extending that implementation. The clients' files are scanned through the java "BasicFileAttributes" interface.

```
import java.sfa.file.Files;
import java.sfa.file.Path;
import java.sfa.file.Paths;
import java.sfa.file.attribute.BasicFileAttributes;
public class FileAttributesSFA{
    public static void main(String[] args) throws Exception {
        String path = "/sfa/input/client1_f1.txt";
        Path file = Paths.get(path);
        BasicFileAttributes attr =
            Files.readAttributes(file, BasicFileAttributes.class);
            System.out.println("creationTime = " + attr.creationTime());
            System.out.println("lastAccessTime = " + attr.lastAccessTime());
            System.out.println("lastModifiedTime = " + attr.lastModifiedTime());
            System.out.println("isDirectory = " + attr.isDirectory()); System.out.println("isOther
            = " + attr.isOther()); System.out.println("isRegularFile = " + attr.isRegularFile());
            System.out.println("isSymbolicLink = " + attr.isSymbolicLink());
            System.out.println("size = " + attr.size());
        }
    }
```

3.1 Merger

This process group sets of files together based on the collected file system attributes, as for many application logic, this can make sense when files are consumedfor processing. For example, in some weather application, it may be interesting to combine files of the same period of each year. In E-learning, files are grouped by disciplines, and so on. The SFA can also combine files based on substring of the file name, which is in many production cases, a hidden attribute of other properties like location, gender, or simply a type of source (station, sensor, satellite…). “See. fig 7”

About the Data

Source	U.S. Geological Survey
Category	GIS, Sensor Data, Satellite Imagery, Natural Resource
Format	GeoTIFF, txt, jpg

- S = Satellite
- PPP = WRS path
- RRR = WRS row
- DDD = Julian day of year
- GSI = Ground station identifier
- VV = Archive version number

Figure 7: AWS dataset example - continuous record of Earth’s land surface as seen from space.

The predefined strategy of the merger is defined in the master configuration file. Common attributessuch as date, size, owner, type and name are assigned coefficients up to the administrator strategy.The coefficients are set by a group of combined files “See. Fig 8"

```
etc root# more sfa_core.conf

#######################
## SFA Attributes
#######################
##Stream Group 1 : customized
group stream_group1
sfa_date_att    1
sfa_type_att    2
sfa_size_att    3
sfa_own_att     4
sfa_cust1_att   5
sfa_substr_att  0

##Stream Group 2 : only by date
group stream_group2
sfa_date_att    1
sfa_type_att    0
sfa_size_att    0
sfa_own_att     0
sfa_substr_att  0
```

Figure 8: SFAcoefficient of attributes per combined queue

To maintain a flexible implementation, one can define a new attribute in a specified section of the master configuration file (exp : sfa_cust1_att in "Fig 8"), then it's possible to use it similarly as the common predefined attributes. The assigned priority is from 0 to 9 while 0 refer to an ignorable attribute. Finally, the list of files is identified in the corresponding queue of the SFA memory. Each file is assigned a value based on the priority of attributes strategy.

3.2 Sorter

This process is tightly coupled to the merger, as the list of files is ready to be consumed. the sorter use the assigned values by the merger to each file, and use a sorting algorithm based on quicksort to generate the ordered final list.

Quicksortis a well-known sorting algorithm[19], that performs the best among its competitors. It's based on divide-and-conquer logic, through the three principals' phases described below, Quicksort is used to sort a random array of numbers, that refer in our case to the generated values for each file by the merger.

Consider an array S[1 .. k]

Phase I: Divide

The input array S[l .. k] is divided in two empty subarrays:

S[l .. m-1] and S[m+1 .. k]. The subarrays are constructed in a way that every left element of S[1 .. m-1] should be smaller or equal to S[m], and every right element of S[m+1 .. k] should be bigger than S[m]. This phase is about choosing the index m, which is called the pivot element.

Phase II: Conquer

Its about sorting each subarray, S[l .. m-1] and S[m+1 .. k] trough recursive calls of quicksort.

Phase III: Combine

One the subarrays are sorted, we combine them together to form the sorted array S[l .. k].

Sorter Algorithm:

Consider a list of numbers S [1 .. k]

```
quicksort(S)

if |S| = 0 then S

else let

p = pick a pivot from S

S1 = s ∈ S | s < p

S2 = s ∈ S | s = p

S3 = s ∈ S | s > p

(S1 ,S3 ) = quicksort(S1 ) || quicksort(S3 )

in

append(S1 ,append(S2 ,S3 ))

end
```

For our implementation, the values assigned by the merger, permit us to use a priority-based selection technique. Where the pivot is not selected randomly in the beginning, then its selected with the highest value. This can help reducing time to choose better pivot.We should emphasize that the pivot elements selected in each subarray created are used only for dividing the elements in the two subarrays, their position is not part of a comparison, but it is a result of the last move of this comparison.

"See. Fig 9"

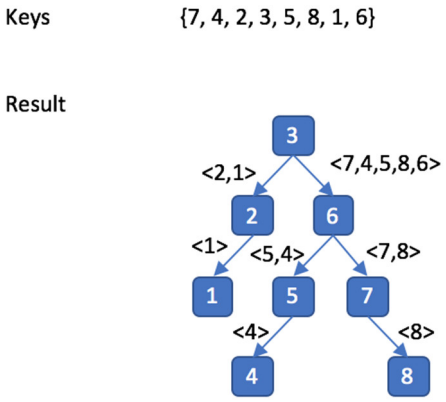


Figure 9: Run example of quicksort

The sorted list of files generated in this phase is stored on the HDFS in MapFile format, which is one of the HDFS supported format, that was introduced to reduce entries in the NameNode memory while dealing with a lot of small files. Moreover, it consists of an indexed SequenceFile to reduce time access compare to sequential read in SequenceFiles. See “Fig 10”.

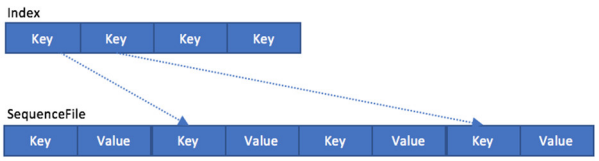


Figure. 10: MapFile File Layout

3.3 Compressor

This process, is not by default activated. In fact, to analyze how the small files are affecting the Hadoop cluster, we import records from the FSimage of the NameNode, which contain a complete state of the HDFS state, then we aggregate data to store it in a reference handled by the database of the SFA, see "Fig. 11".

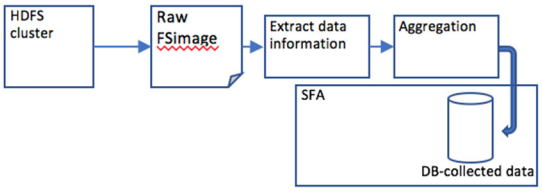


Figure. 11: SFA reference steps

This reference is completely independent of the functionality of our SFA sever, and can be built on a new server if the performance is impacted. Though, it allow us to get some very useful information and reports about the small files in our cluster, such as :

- Files < 512 Kb)
- Files < 512 Kb (Grouped by Path)
- Files < 512 Kb (Grouped by Owner)
- Files < 512 Kb (Grouped by Date)
- Number of blocks with a used space < 60%
- Most called files per MapFile (hot small files)
- Files being called in more than one month (per user, per path, per MapFile)
- …

The compressor process, is used to compress MapFiles values when relevant, scheduling this feature is not recommended but it remain possible in this implementation based on MapFiles age if needed.

3.4 File Prefetching and Caching

Prefetching and caching technique has been implemented in this study, to improve MapReduce jobs performance when small files are processed. In fact, prefetching can reduce the disk I/O overhead, and improve the access time as data can be brought to the cache before being requested. When a small file is requested, the metadata for the MapFile are pulled from the NameNode and the DataNodes that holds the related blocks is identified. Though, if the metadata of that MapFile exist in the local memory of the SFA server, no request is sent to the NameNode. Therefore, it becomes efficient to read small files with less requests sent to the NameNode. Also, on the DataNode level, a small file request, is usually combined with the MapFile that contain it, that specific format provide an index to determine the location of each small file through its key and then its offset and length. Once the small file is located, the MapFile index is cached in the SFA memory, and the specifics requested small files are cached in the DataNode memory. This can help accessing contained files directly in the next subsequent calls. To avoid memory overhead on the DataNodes, LRU algorithm was used to keep most called objects and manage future allocations

efficiently [20]. Caching the specified objects in the DataNodes are achieved through mlock() and mmap(), because storing data off-heap will not affect garbage collection when the amount of cached data is large.

4 Experimental Evaluation

The experimental test platform is in this paper is achieved on an implemented cluster of one NameNode, 4 Datanodes, and the SFA server. The NameNode is a server of 3.10 GHz clock speed, 16GB of RAM and a gigabit Ethernet NIC. Each DataNode offer a 500GB Hard Disk, and they are deployed on Ubuntu 14.04. The SFA server is built on 3.10 GHz clock speed, 16GB of RAM, and deployed on Ubuntu 14.04.The replication factor is kept as the default value 3, and the block size of HDFS is chosen as 64Mb. The experimental datasets are basically standard auto-generated.We use Hadoop version 2.6.2 and java version 1.8.0_65.

To evaluate our approach, the same datasets are stored in three different ways. Initially, the datasets are uploaded directly on HDFS without grouping in any format. Then the datasets are uploaded into HDFS using Hadoop Archive (HAR format).

Finally, we stored files through the SFA server to compare the result of the previous strategies with our approach. The evaluation was based on memory usage and time-cost for writing and reading small files.In the HAR case, original files are deleted at the end of HAR merging.

4.1 Comparison of the NameNode Memory Usage

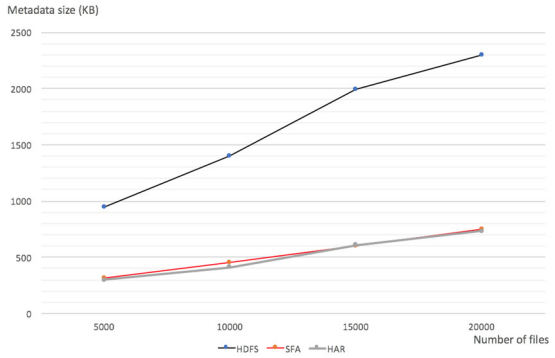


Figure. 12: NameNode memory usage

According to “Fig. 12”, when we store small files through SFA, or in HAR format, The NameNode memory consumption is too low due to the reduced entries of metadata, as each list of files in MapFile or HAR file require only the related metadata for the whole merged file.

4.2 Comparison ofthe required time for storage

To compare the required time for storage, we uploaded a set of datasets, from 1000to 20 000 small files, in 3 ways as described for memory usage evaluation.

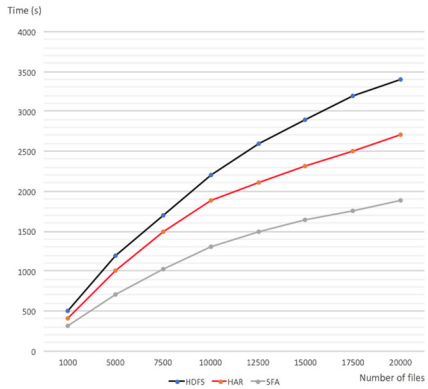


Figure. 13 Comparison of the required time for storage

According to"Fig. 13", the SFA greatly outperforms native HDFS and the HAR. This is because combing clients' files and attributes aggregation allow better block allocations, and reduces the NameNode requests for each client's writing request.

4.3 Comparison of the random and sequentialreading time of small files

To evaluate the reading time, we selected randomly different number of small files from each previous upload case (HDFS, HAR and SFA). We downloaded randomly 200, 500, 1000, 1500, 2000, 4000, 6000 and 8000 files from the total amount (20 000 files) in each case. Respectively we evaluate the required time for each download operation.

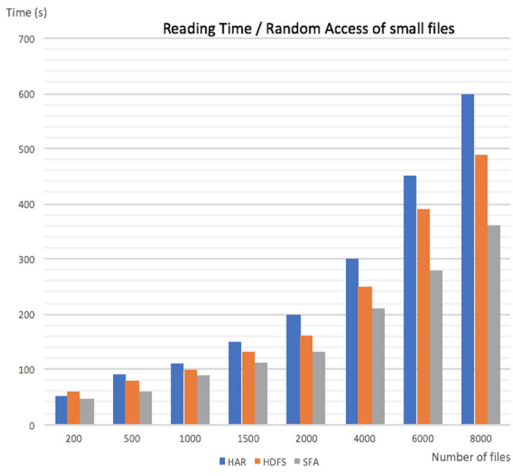


Figure. 14 Comparison of the reading time of random small files

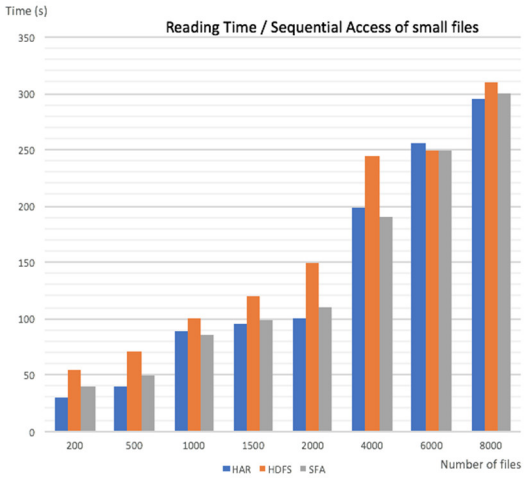


Figure. 15 Comparison of the reading time of sequential small files

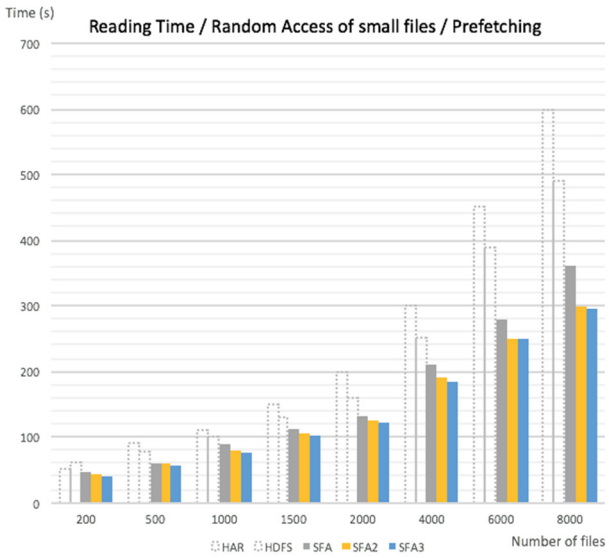


Figure. 16 Comparison of the reading time of random small files / 3 times

According to"Fig. 14" and "Fig. 15",when reading random selected files, SFA can reduce the access time by 24% and 13% compared to HDFS and HAR respectively. When reading sequential lists of files, SFA outperforms HDFS but it remains a bit slower than HAR. The reason is that the increase of the number of files can increase the number of seek operations on DataNodes through the index structure of MapFiles, which can result in the higher reading latency. For SFA, reading random files results in better access efficiency than reading sequential lists of files. Moreover, according to "Fig. 16",the random-access time can be improved after the second and third call of the same sets of small files, which proves clearly the efficiency of prefetching and caching mechanism on the DataNodes and the SFA sever.

5 Conclsion and Future Works

In this paper, we focus on improving the cluster performance while processing small files. The SFA server has been used to combine clients' small files into MapFiles. This technique has been improved in the global approach, which consist of grouping elements in specific order. Sorting small files per categories, up to the application logic, is a promising concept that can improve the reading time access. In fact, the memory usage has been addressed as the main problem in many researches, while in this study, we consider it as only one part of the small files problem equation. Our approach addresses the aspect of how the distribution of small files in a specific format can influence the cluster performance. Different strategies are now adopted in Hadoop to solve the small files problem, but there is a lack of standardization, as most of the solutions remain useful in specific environmentsbut not in others. Offering a system to analyze different aspects of the small files problem can help organizations to understand better the real factors that control the impact of their datasets.

REFERENCES

[1] B. White, T. Yeh, J. Lin, and L. Davis, "Web-scale computer vision using mapreduce for multimedia data mining," in Proceedings of the Tenth International Workshop on Multimedia Data Mining. ACM, 2010, p. 9.

[2] K. Wiley, A. Connolly, J. Gardner, S. Krughoff, M. Balazinska, B. Howe, Y. Kwon, and Y. Bu, “Astronomy in the cloud: using mapreduce for image co-addition,” Astronomy, vol. 123, no. 901, pp. 366–380, 2011.

[3] W. Fang, V. Sheng, X. Wen, and W. Pan, “Meteorological data analysis using mapreduce,” The Scientific World Journal, vol. 2014, 2014.

[4] F. Wang and M. Liao, “A map-reduce based fast speaker recognition,” in Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on. IEEE, 2013, pp. 1–5.

[5] K. P. Ajay, K. C. Gouda, H. R. Nagesh, “A Study for Handelling of High-Performance Climate Data using Hadoop, Proceedings of the International Conference, pp: 197-202, April 2015.

[6] D. Q. Duffy, J. L. Schnase, J. H. Thompson, S. M. Freeman, and T. L. Clune, “Preliminary evaluation of mapreduce for high-performance climate data analysis,” 2012.

[7] C. Shen, W. Lu, J. Wu, and B. Wei, “A digital library architecture supporting massive small files and efficient replica maintenance,” in Proceedings of the 10th Annual Joint Conference on Digital Libraries (JCDL '10), pp. 391–392, June 2010

[8] A. Thusoo, Z. Shao, S. Anthony, D. Borthakur, N. Jain, J. Sen Sarma, R. Murthy, and H. Liu, “Data warehousing and analytics infrastructure at facebook,” in Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010, pp. 1013–1020.

[9] J. K. Bonfield and R. Staden, “ZTR: A new format for DNA sequence trace data”, Bioinformatics, vol. 18, no. 1, (2002), pp. 3–10.

[10] T. Nukarapu, B. Tang, L. Wang, and S. Lu, “Data replication in data intensive scientific applications with performance guarantee,” IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 8, pp. 1299–1306, 2011.

[11] Y. Gao and S. Zheng, "A Metadata Access Strategy of Learning Resources Based on HDFS," in proceeding International Conference on Image Analysis and Signal Processing (IASP), pp. 620—622, 2011.

[12] J. Xie, S. Yin, et al. “Improving MapReduce performance through data placement in heterogeneous Hadoop clusters”, In 2010 IEEE International Symposium on Parallel & Distributed

[13] G. Mackey; S. Sehrish; J. Wang. Improving metadata management for small files in HDFS. IEEE International Conference on Cluster Computing and Workshops (CLUSTR). 2009. pp.1-4.

[14] C. Vorapongkitipun; N. Nupairoj. Improving performance of small-file accessing in Hadoop. IEEE International Conference on Computer Science and Software Engineering (JCSSE). 2014. pp.200-205.

[15] Patel A, Mehta M A. A novel approach for efficient handling of small files in HDFS, 2015 IEEE International Advance Computing Conference (IACC), pp. 1258-1262.

[16] Y. Zhang; D. Liu. Improving the Efficiency of Storing for Small Files in HDFS. International Conference on Computer Science & Service System (CSSS). 2012. pp.2239-2242

[17] D. Dev; R. Patgiri. HAR+: Archive and metadata distribution! Why not both?. IEEE International Conference on Computer Communication and Informatics (ICCCI). 2015. pp.1-6.

[18] P. Gohil; B. Panchal; J. S. Dhobi. A novel approach to improve the performance of Hadoop in handling of small files. International Conference on Electrical, Computer and Communication Technologies (ICECCT). 2015. pp.1-5.

[19] Hoare CAR. Quicksort[J]. The Computer Journal, 1962, 5(1): 10-16

[20] E.J. O'Neil, P.E. O'Neil, and G. Weikum, "An Opti- mality Proof of the LRU-K Page Replacement Algorithm," J. ACM, vol. 46, no. 1, 1999, pp. 92-112

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Implementation of the Flexible "Private - Public" Cloud Solution based on OpenStack

M. Elmahouti, O. Achandair, S. Khoulji, M.L. Kerkeb
Laboratory ISERG, Abdelmalek Essaadi University, Tétouan, Morocco
mouradelmahouti@gmail.com, o.achandair@gmail.com, khouljisamira@gmail.com,
kerkebml@gmail.com, acmlis.conference@gmail.com

ABSTRACT

Cloud computing is a model that facilitates access to and manipulation of resources on demand. It is a technology that is unique today to meet the needs and demands of customers by guaranteeing a high quality of service rendered. This new model provides convenience to reorganize the current revolt in the information technology industry by ensuring cost effective and less costly solutions to meet the constraints of technical capabilities and their extension.

This article will present the technical implementation of a flexible "private - public" cloud computing, based on the Openstack solution, to ensure the business needs in terms of performance, and a response time of treatments tailored to customer demand with a Availability.

This approach, followed by a "private-public" flexible cloud, will be able to communicate two clouds, the Cloud A, which includes the entire physical infrastructure of the company, while cloud B will be provided by a service provider that does not Will be called once the configurable load threshold is exceeded on the cloud A, and as soon as the resources on the private cloud A are released, the instances migrated to the Cloud B will be again remigrated to the Cloud A to minimize even the times Allocation.

Keywords—Cloud computing, Private cloud, Public cloud Information resources, Information security, OpenStack

1 Introduction

Today, cloud computing technology has marked an intrinsic transformation in improving the quality of services rendered, minimizing response times and processing on exponential quantities of dynamic data.

It is not easy today to master the growing difficulties for the proper functioning of the different architectures, which have to adapt with the required load to meet a certain level of performance within the company, this new concept of Cloud computing offers fascinating proposals to companies to choose the optimal configuration to outsource some or the entire infrastructure.

The article addresses the different concerns of information systems managers who want to join this new cloud computing technology and have slowed down because of the significant investment in their current infrastructure by offering a smart and flexible solution based on free software "OPENSTACK", which

DOI: 10.14738/tmlai.54.3334
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3334

consists of creating two clouds, a private cloud that supports the company's current platform, and a public cloud that will be auxiliary once the load threshold exceeds the capabilities of the internal platform.

The article is considered a continuation of the previous article [19], or we will detail the technical implementation of this flexible private cloud solution "private - Public", giving the necessary screenshots.

2 Proposed Solution: Private-Public Cloud Model Based On Openstack

To encourage companies to adopt this new cloud computing technology, minimizing budgets in the future and taking advantage of the state of the existing and existing platforms, our proposal is to create two models "Private : Cloud A"And" public : Cloud B "cloud based on a free solution that is the OpenStack[1].

Our approach is to create a "private - public" cloud: a private cloud (Cloud A) that supports the existing platform, and once the number of requests exceeds the capacity of the private platform, these requests will be refused internally by routing them to the Cloud to satisfy them by the public cloud (Cloud B) by paying just the rental costs, And once the internal resources are released, automatically some of the applications running on the public cloud will be supported again by the private cloud by minimizing rental costs[2].

2.1 Architecture proposed for the cloud "Private - Public"

The “Figure 1”shows the mechanism to be used by proposing the three algorithms to follow in order to pass the two-way passage between the two clouds, private and public [3], to ensure this proposition:

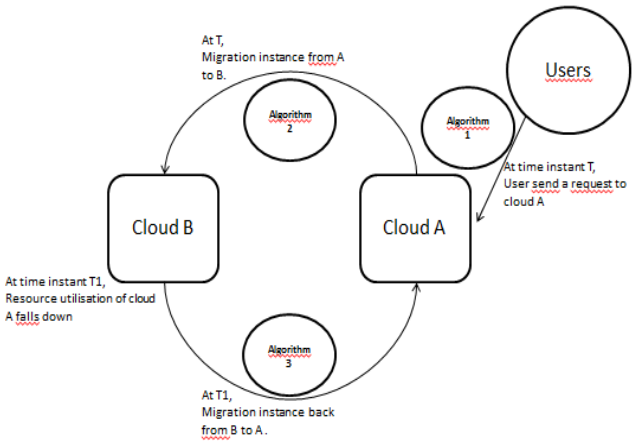


Figure 1: Cloud integration

2.2 Presentation of the three algorithms used:

The main brick of Openstack is Nova. Its purpose is to manage the resources of calculation of the infrastructure used, using the command NOVA BOOT to attack our platform [4], with its syntax below:

nova boot --image *imageID* --flavor *flavorID* --nic net-id=*nicID*

imageID **:**To select the operating system.

net-id=*nicID* **:** To tell the cloud which subnet is used for an instance.

flavorID **:**Reflects memory, disk and virtual process requirements.

To check the actual load on the cloud A, an implementation of the load analysis function is done, the functioning of the "analyze load" function will be explained in the algorithm 1 below[16].

Once the "analyze load" function is called, all information from all compute nodes of a controller on cloud A is received, if the load on the A cloud is below the configurable threshold, the instance will be created On cloud A, but if the load is greater than the threshold, the instance will be created on the cloud B[5].

Algorithm 1 Analyzing the load

Input Parameter:
flavorid: flavor id of requested instance

Output Parameter:
val: 0 if cloud is overloaded, 1 otherwise.

```
procedureAnalyzing the load (f lavor id)
Oracle database connection is made to the nova
database.
Extract user's requirements using flavor id from the
instance types table.
totalmemory for all compute and total disk for all
compute is initialized to 0.
freememory for all compute and free disk for all
compute is initialized to 0.
for each compute node c do
Find the total memory and the free memory of c
from the compute nodes table.
Find the total disk space and the free disk space of
c from the compute nodes table.
totalmemory for all compute =total memory for all
compute + total memory of c.
 total disk for all compute =total disk for all c-
ompute + total disk of c.
 freememory for all compute =free memory for all c-
ompute + free memory of c.
free disk for all compute =free disk for all co-
mpute + free disk of c.
end for
if (total disk for all compute * disk threshold
percentage>= users disk requirement + (to-
tal disk for all compute-free disk for all
compute) then
if (total memory for all compute * ram threshold
percentage) >= users memory requirement +
(total memory for all compute-free memory for
 all compute) then
Return 1
end if
end if
Return 0
end procedure
```

Note that setting up an instance on another cloud requires some information on the Cloud B: Image-ID, Network-ID, which is used by algorithm 2 to position the instance on the cloud B[6].

Algorithm 2 Position the instance on the cloud B

Input parameter:
Network id: Subnet to start instance. Image id: OS launching an instance.
Tasting Id: Specify required resources
Instance Name: The name of the instance to be launched.
Procedure INSTANCE positioning ()
When running the nova boot command, the function creation ()
calls up the load to check the load
If Not exceeding threshold then
The instance is created on cloud A.
Other

It calls the createInstance () function. // Creating the instance on the cloud B
CreateInstance () uses the new startup parameters
We call the function creation () and extract the
Id of the image for the cloud B.
It Execute the NOVA command with the new parameters.
This nova boot command is executed
On the B cloud using the SSH connection.
All instance information is stored in a file, retrieved later, and stored again in the array of migrating instances to the NOVA database
End if
End of procedure.

After each migration to the cloud B, the load on the cloud A is verified, once the load is below the configurable threshold, a remigration is made again to the cloud A, the instances that will be migrated back to the cloud A follow A FIFO order, algorithm 3 below explains the function of the remigration operation [7]:

Algorithm 3 Remigration

Input Parameter:
Migratedinstancestable: Table regrouping the instances of remigration
procedure REMIGRATION( )
The Analyzing the load() for remigration is Performed at
Of the parameterized intervals.
if the resource utilization falls below a configurable
low threshold then
do
It finds the instance to be remigrated from the
migratedinstances table using the FIFO : rule.
It launches a new instance on cloud A using
the information stored in the migrated instances table.
It copies the disk image of the migrated
instance from the cloud B to the newly launched instance on the cloud A, to: restore the current state.
It finished the remigrated instance from thecloud B to free resources.
It cleans the entry of the table of instances that have been migrated.
whileResource usage is below the migration threshold and the migrated instance table is not empty
end if
end procedure

The communication between the two clouds (Cloud A and Cloud B) is secured with the use of an SSH public key, as noted in algorithm 2, [8].

3 The OpenStack SOLUTION

3.1 OpenStack Architecture

OpenStack is a set of open source software that enables deployment and manipulation of cloud computing infrastructures. This technology has a modular architecture that consists of several correlated sub-projects (Nova, Swift, Glance .....), see “Figure 2”, whose purpose is to control all the resources of virtual machines, such as computing power, storage or network [9].

The Openstack project is supported by the Openstack Foundation, a non-commercial organization whose goal is to promote the Openstack project by providing support and support to the entire OpenStack community. [10]

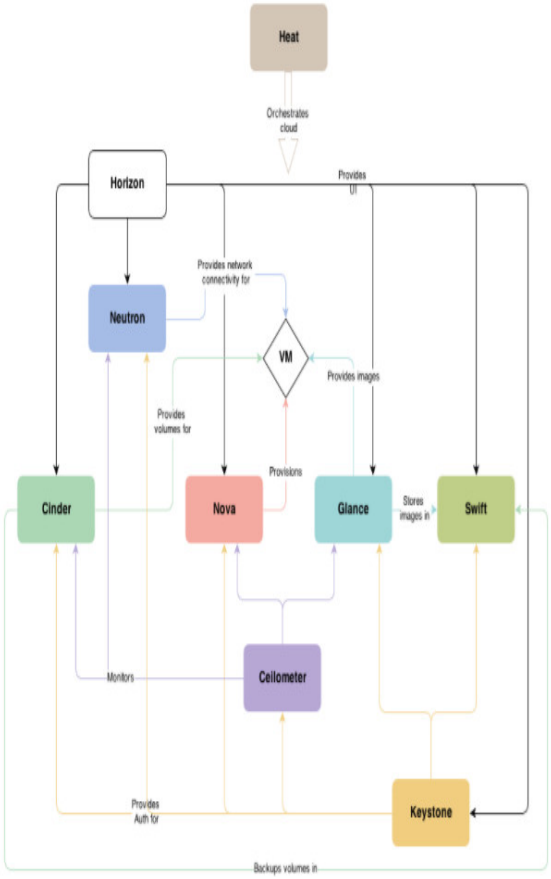


Figure 2: OpenStack Architecture and Components

3.2 The components of OpenStack

- Calculation / Nova: Nova is one of the main components of Openstack. Its role is to manage infrastructure calculation resources [11].

Nova's architecture is designed to evolve horizontally by adding hardware. Among the strengths of Nova is that it works with non-specialized hardware, which makes it possible to reuse existing servers for example [12].

- Object Storage / Swift: It is a dedicated system for storing redundant and scalable data, files are written to multiple hard disks spread across multiple servers, responsible for replication and data integrity within the cluster [13].

  The Swift cluster evolves horizontally by adding new servers, once a server or disk fails, Swift takes care of replicating its content from active nodes.

  As Swift is based on an application logic, it allows the use of inexpensive and non-specialized equipment [14].
- Block storage / Cinder:

  This is the OpenStack block storage service; its role is to provide block-based persistent devices to OpenStack instances.

  It manages the creation, attachment, and detach operations of these devices on servers.

  This bulk storage mode is used for high-performance scenarios such as database storage, but also to provide the server with low-level access to the storage device [15].
- Network / Neutron:

  The Neutron service allows you to manage and manipulate networks and IP addressing within OpenStack, allowing users to create their own networks by controlling traffic and connecting their instances to one or more networks.

  Neutron also provides different types of network deployment based on the target infrastructure.
- Dashboard / Horizon: This is a WEB application that allows users and administrators to have a dashboard to manage their clouds through a GUI [16].
- Identity service / Keystone: This service provides a central directory containing the list of services and the list of OpenStack users as well as the roles and permissions.

  Within OpenStack, Keystone is used by all users and services to authenticate with each other.
- Image service / Glance: This is the OpenStack image service, which allows the discovery, sending and distribution of disk images to instances. The glance service also allows you to store backups of these disks. Glance can store these disk images in several ways: in a folder on server, but also through the object storage service of OpenStack or in decentralized storage [17].
- Telemetry / Ceilometer: It is the service that collects different metrics on cloud usage. For example, the number of instances launched in a project and for how long.

  These metrics can be used to provide information needed for a billing system, for example. These metrics are also used in applications or other Openstack components to define actions based on certain thresholds as with the orchestration component.
- Orchestration / Heat: It is the component that ensures the orchestration of OpenStack, it allows to describe an infrastructure in the form of models. It can also use the metrics provided by Ceilmeter to decide to create additional instances according to the load of an application for example [18].

4 Implementation of the Solution: Private-Public Cloud Model Based on Openstack

To implement the proposed solution of creating two clouds, a private cloud A that supports the company's internal infrastructure, and a public cloud B at our service provider, which will be called as an ancillary platform, Once the load threshold is exceeded in the cloud A, the newly launched instances will be migrated to the cloud B for execution, once the load threshold on the cloud A arrives at its configured value, executing instances and Loaded by the public cloud B will be migrated back to the private cloud A.

An automatic load metering is initiated every 20 seconds to evaluate the load threshold in the cloud A to make the right decision on new instances arriving at cloud A and those running on public cloud B to ensure Either a migration to the B cloud when the threshold is exceeded or a remigration to the private cloud A.

In order to technically ensure this solution, two DELL Poweredge 2950 servers were used (one server is bought by our company, the other is leased since it is provided by our service provider, the same range was required to ensure efficient compatibility Between services), each having a memory of 1.8 TB, which acts as one of the compute nodes.

100 machines were assigned which function as calculation nodes for the two clouds, respectively. Knowing that the two clouds belong to a single subnet.

The OpenStack solution was installed on both servers, an installation based on Ubuntu 16.04, and then two clouds, a private cloud on the first server, all the internal servers in the company, and a second public cloud at the provider that will be used after the configurable threshold is exceeded.

The private cloud A is used by all production departments to execute their instances, they can create instances with different dimensions of 20 GB, 40 GB or 80 GB. Once 80% of the disks are allocated and 75% Of memory is busy, Cloud A is said to be overloaded, and all other requests will be satisfied by the cloud B.

Since both clouds belong to the same subnet, migrating an instance of cloud A to cloud B takes almost 17 seconds (for example, for an instance of 20GB). This time varies depending on the size of the instance.

The “Figure 3” shows the initial state of the public cloud B which initially contains four instances.

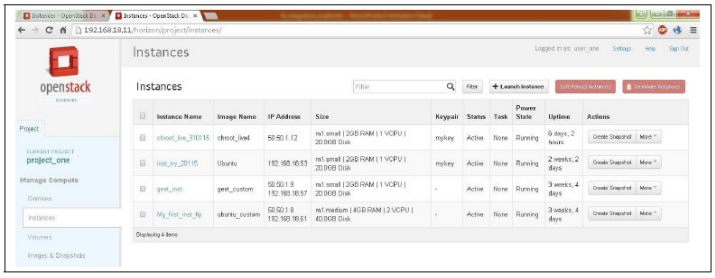


Figure 3: Initial Public Cloud

The “Figure 4” depicts the placement process of an instance; the terminal at the bottom of the “Figure 4” shows the execution of the new start command on the private cloud A.

The dashboard on the right side of the figure shows multiple instances that run on the private cloud.

The private cloud is overloaded because of these instances, so an instance named inst6 will be created on public cloud B as shown in the dashboard on the left side of the figure.

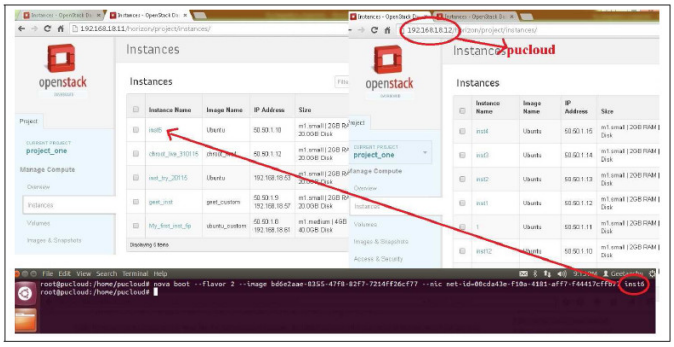


Figure 4: Placement of the instance on public cloud B.

The "Figure 5" shows that once the private cloud is overloaded with reaching the configurable threshold, the creation of the four instances (from inst6 to inst9) is supported by the public cloud.

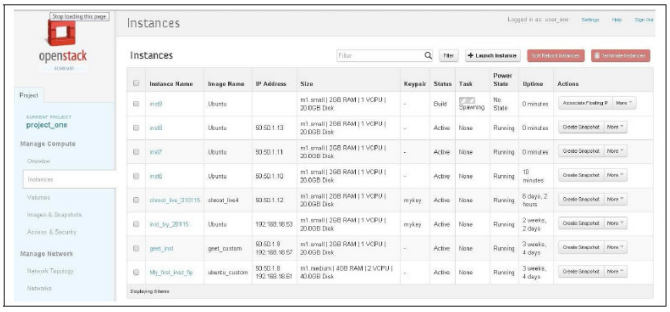


Figure 5: The public cloud after overloading condition of private cloud.

Some instances from private cloud are terminated to reduce the load on it, till it reaches the configurable low threshold as show in the "Figure 6".

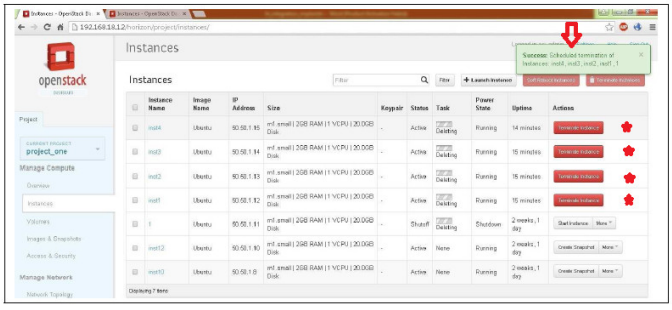


Figure 6: Reducing load on the private cloud.

Once the load is reduced on the private cloud, and we arrive below the reconfigurable threshold, the migration process is triggered.

The “Figure 7” shows the remigration of the instance inst5.

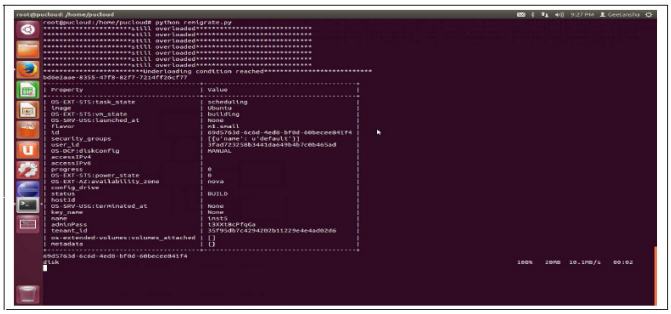


Figure 7: The remigration process.

The “Figure 8” and “Figure 9” show the state of both private and public clouds after re-migration.

Once the load is reduced on the private cloud, four instances (De inst6 to inst9) have been remigrated to the private cloud by releasing the public cloud.

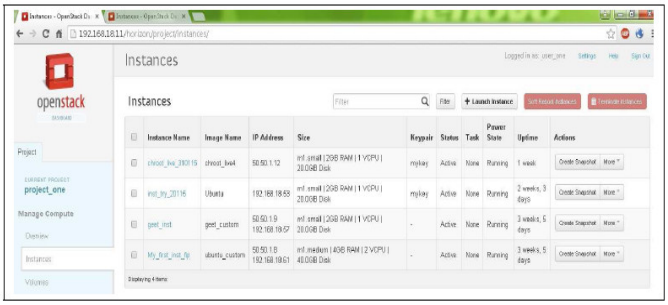


Figure 8: The public cloud after remigration.

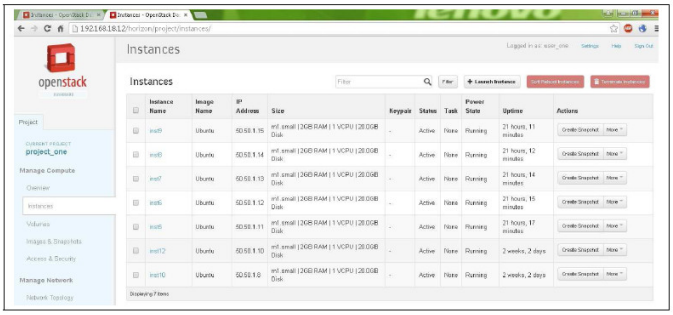


Figure 9: The private cloud after remigration.

5 Conclusion

In this paper, we proposed an effective solution to accompany the change in information systems management, especially to encourage the adoption of new cloud computing technology based on the free OpenStack solution to create a private cloud Which consolidated the entire enterprise infrastructure and another public cloud at the service provider that is called whenever internal capabilities are exhausted and the only configurable is outdated. This solution has been definitively deployed and has shown these added values in terms of performance, quality of service rendered and associated costs.

The current implementation is considered the first pilot site, the next step will be to generalize the same architecture for all the subsidiaries.

ACKNOWLEDGMENT

This work was carried out with the support of our company, for the implementation of the project.

We thank all members of the project team who contributed to the success of this proposal.

REFERENCES

[1] Haifeng Li, Huachun Zhou, Hongke Zhang, Bohao Feng, "EmuStack: An OpenStack-Based DTN Network Emulation Platform", 2016 International Conference on Networking and Network Applications (NaNA), vol. 00, no. , pp. 387-392, 2016, doi:10.1109/NaNA.2016.24

[2] Dinkar Sitaram, Sudheendra Harwalkar, K.V. Suchith Kumar, "Standards Based Integration of Intercloud for Federation with OpenStack", 2016 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), vol. 00, no. , pp. 113-118, 2016, doi:10.1109/CCEM.2016.028

[3] Pooya Musavi, Bram Adams, Foutse Khomh, "Experience Report: An Empirical Study of API Failures in OpenStack Cloud Environments", 2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE), vol. 00, no. , pp. 424-434, 2016, doi:10.1109/ISSRE.2016.42

[4] M. Abhishek Sharma, Monica O. Joshi, "Openstack Ceilometer Data Analytics & Predictions", 2016 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), vol. 00, no. , pp. 182-183, 2016, doi:10.1109/CCEM.2016.045

[5] Ali Babar, Ben Ramsey, "Tutorial: Building Secure and Scalable Private Cloud Infrastructure with Open Stack", 2015 IEEE 19th International Enterprise Distributed Object Computing Workshop (EDOCW), vol. 00, no. , pp. 166, 2015, doi:10.1109/EDOCW.2015.23

[6] Dario Bruneo, Francesco Longo, Giovanni Merlino, Nicola Peditto, Carmelo Romeo, Fabio Verboso, Antonio Puliafito, "A Modular Approach to Collaborative Development in an OpenStack Testbed", 2015 IEEE Fourth Symposium on Network Cloud Computing and Applications, vol. 00, no. , pp. 7-14, 2015, doi:10.1109/NCCA.2015.12

[7] Vojtech Cima, Bruno Grazioli, SeÃ¡n Murphy, Thomas Michael Bohnert, "Adding energy efficiency to Openstack", 2015 Sustainable Internet and ICT for Sustainability (SustainIT), vol. 00, no. , pp. 1-8, 2015, doi:10.1109/SustainIT.2015.7101358

[8] D. Thomas Erl, D. Zaigham Mahmood, D. Ricardo Puttini, “Cloud Computing: Concepts, Technology & Architecture” September 2015

[9] Yun Zhang, Farhan Patwa, Ravi Sandhu, Bo Tang, "Hierarchical Secure Information and Resource Sharing in OpenStack Community Cloud", 2015 IEEE International Conference on Information Reuse and Integration (IRI), vol. 00, no. , pp. 419-426, 2015, doi:10.1109/IRI.2015.71

[10] Robayet Nasim, Andreas J. Kassler, "Deploying OpenStack: Virtual Infrastructure or Dedicated Hardware", , vol. 00, no. , pp. 84-89, 2014, doi:10.1109/COMPSACW.2014.18

[11] David Bernstein, "Cloud Foundry Aims to Become the OpenStack of PaaS", IEEE Transaction on Cloud Computing, vol. 1, no. , pp. 57-60, July 2014, doi:10.1109/MCC.2014.32

[12] Juan Angel Lorenzo del Castillo, Kate Mallichan, Yahya Al-Hazmi, "OpenStack Federation in Experimentation Multi-cloud Testbeds", , vol. 02, no. , pp. 51-56, 2013, doi:10.1109/CloudCom.2013.103

[13] Yang Wang, Xiaomin Wang, Jianyong Chen, "On-Demand Security Architecture for Cloud Computing", IEEE Computer, vol. 45, no. , pp. 73-78, July 2012, doi:10.1109/MC.2012.120

[14] V.KRISHNA REDDY1 , Dr. L.S.S.REDDY, "Security Architecture of Cloud Computing",International Journal of Engineering Science and Technology (IJEST), ISSN : 0975-5462 Vol. 3 No. 9 September 2011 pp.7149-7155.

[15] A Beloglazov, J Abawajy, R Buyya - Future generation computer systems, Future Generation Computer Systems, Elsevier, Volume 28, Issue 5, May 2012, Pages 755–768

[16] Ian Foster, Rubén S. Montero, Borja Sotomayor, Ignacio M. Llorente, "Virtual Infrastructure Management in Private and Hybrid Clouds", IEEE Internet Computing, vol. 13, no. , pp. 14-22, September/October 2009, doi:10.1109/MIC.2009.119

[17] Wen-Hwa Liao, Shuo-Chun Su, "A Dynamic VPN Architecture for Private Cloud Computing", 2011 IEEE 4th International Conference on Utility and Cloud Computing (UCC 2011), vol. 00, no. , pp. 409-414, 2011, doi:10.1109/UCC.2011.68

[18] Sumit Goyal, "Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review", I.J. Computer Network and Information Security, 2014, 3, 20-29.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

An Efficient Sky Detection Algorithm from Fisheye Image Based on Region Classification and Segment Analysis

[1]Houssam Bourr, [1]Youssef El Merabet, [1]Rochdi Messoussi, [1]Ibtissam Benmiloud, [2]Yassine Ruichek,
[1]LASTID Laboratory, Department of Physics,
Faculty of Sciences, Ibn Tofail University
Kenitra, Morocco
[2]Le2i FRE2005, CNRS, Arts et Metiers, Univ. Bourgogne Franche-Comte
Belfort,France
bourr.houssam@gmail.com; y.el-merabet@univ-ibntofail.ac.ma;
yassine.ruichek@utbm.fr; messousi@gmail.com

ABSTRACT

In this paper, an efficient approach for automatic and accurate sky region detection from fisheye images is proposed. The proposed approach starts by segmenting the acquired image into regions using Statistical Region Merging method. After that, the segmented regions are characterized using local RGB color descriptor using image quantization. The next step consists of classifying the characterized regions into sky and non-sky regions by using maximal similarity based region classification through Hellinger kernel-based distance. In order to improve the obtained region classification results, a segment analysis based technique using Line Segment Detector is proposed. Experimental results prove the robustness and performance of the proposed procedure.

Region classification, RGB color descriptor, Segment analysis, LSD, Hellinger kernel-based distance.

1 Introduction

Sky region detection is necessary for many applications and it is one of the most significant subject matter commonly seen in outdoorimages. In the application of autonomous ground robot navigation, a robot should understand its surrounding environment by visual sensors; A precise sky detection can improve efficiency of road detection for independent ground robot. Sky detection can improve accuracy of GNSS based localization [3, 5, 9], can be used for scene classification, and is often employed to achieve more effective content-based image retrieval [6]. Most existing sky region detection algorithms are mainly based on color priors [1, 2, 11]. In the method proposed in [1], both image simplification and classification are employed. First, the acquired image is simplified using a geodesic reconstruction [10] with an optimal contrast parameter. Second, the two classes (sky and not-sky) are classified from the simplified image. For that, a set of supervised (Bayes, KNN and SVM) and unsupervised (Fuzzy C-means, KMlocal,Statistical Region Merging and Fisher) clustering algorithms are compared with the purpose todefine the best classifier in terms of good classification rate. The method proposed by El merabet et al. [2] is constituted of four major steps: (I) A simplification step that consists in simplifying the image with an appropriate

DOI: 10.14738/tmlai.54.3335
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3335

couple of colorimetric invariant and exponential transform. (II) The second step segments the simplified image in different regions of interest using Statistical Region Merging segmentation method. (III) The third step, the segmented regions are represented with a number of local color image region descriptors. (IV) The fourth step applies a supervised MSRC (Maximal Similarity Based Region Classification) method by using Bhattacharyya coefficient-based distance to classify the characterized regions into sky and non-sky regions.

Unfortunately, this color information based sky region detection approaches are often accompanied with false positives especially for some challenging cases like presence of buildings with the same color of sky (generally blue, white), or windows with mirrors or even mirrored buildings that reflect the color of sky. The purpose of this research is the development of an algorithm for sky detection from fisheye images with more accuracy, which should effectively deal with the above mentioned drawbacks. The method we propose is composed of two main phases: 1) an image-content classification method is used in order to classify the regions of the image into sky and non-sky and 2) an LSD [13] (Line Segment Detector) based segment analysis technique is performed in order to refine the obtained classification results.

The paper is organized as follows: Section 2 presents the image-content classification method employed in this research. Section 3 presents the LSD based segment analysis technique introduced to refine the classification results obtained with the image-content classification method presented in Section 2. Experimental results are shown in Section 4. Conclusions and future works are derived in Section 5.

2 Image-Content Classification Method

In this work, in order to obtain the preliminary classification results of the acquired fisheye images into sky and non-sky regions, we use the method introduced in [4]. This method is composed of several steps (cf. Figure 1): 1/ image simplification, 2/ image segmentation, 3/ region features extraction and 4/ region classification. these steps will be detailed in next sections.

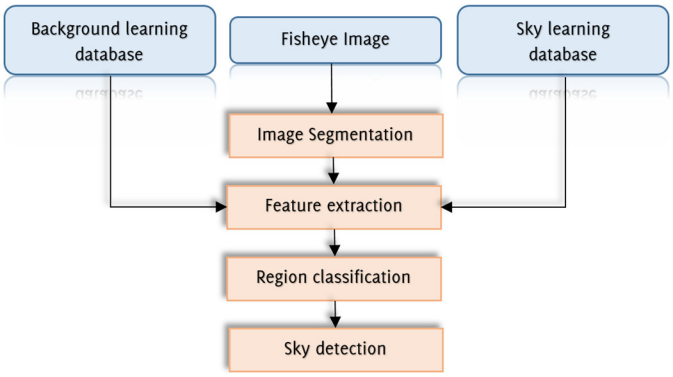


Figure 1: Flowchart of the image-content classification method.

2.1 Initial segmentation using Statistical Region Merging

As pointed out previously, the first step of the framework introduced in previous work of the authors [4] consists in partitioning acquired images into homogeneous regions with the same properties. Obviously, poor segmentation results could potentially impact the system reliability, i.e., the quality of classification results provided by the proposed procedure is strongly dependent on the segmentation results. In this work, in order to obtain a preliminary fisheye image segmentation, we have, like [4], used SRM (Statistical Region Merging) algorithm [8] that seems to be more adapted when considering the objectives of our

application. SRM algorithm presents several advantages: 1/ it allows defining a hierarchy of partitions; 2/ it runs in linear-time by using bucket sorting algorithm while traversing the region adjacency graph (RAG); 3/ it dispenses dynamical maintenance of RAG; and 4/ it does not only consider spectral, shape and scale information, but also has the ability to cope with significant noise corruption and handle occlusions. Figure 2 illustrates an example of segmentation result via SRM algorithm.

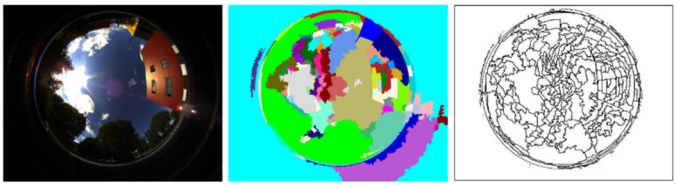


Figure 2: Segmentation result using Statistical Region Merging (SRM) method. From left to right: initial image, its SRM segmentation result and corresponding region map.

2.2 Region representation

The main goal of this research consists in identifying the regions corresponding to sky in fisheye images. For this purpose, we need first to characterize the segmented regions, obtained by using SRM algorithm, with a suitable descriptor. Œe extraction of this local image region descriptor, used as input to the region classification algorithm, permits to provide a global region representation with high level description than raw image pixels allowing thus, to discriminate robustly between the different regions in the treated images. In the present study, we have used RGB color histogram which is implemented as follows: each RGB color channel is first uniformly quantized into l levels, after that, the color histogram of each segmented region is produced in the feature space of z = l × l × l bins. Let I be an image containing N pixels quantized in z=16×16×16=4096 color bins, the RGB color histogram of a segmented region $\mathcal{R}$ is represented as.

$$h^{RGB}(\mathcal{R}) = [h^1_{\mathcal{R}}, h^2_{\mathcal{R}}, \dots, h^z_{\mathcal{R}}] \tag{1}$$

Where

$$\mathbf{h}^i_{\mathcal{R}} = \sqrt{(\frac{\sum_{j=1}^{N} p_{i|j}}{\tau})}; \mathrm{j} \in \mathcal{R} \ \ and \ \ 0 \leqslant \mathrm{i} \leqslant \mathrm{z}. \tag{2}$$

$\mathbf{h}^i_{\mathcal{R}}$ is the *i*th normalized histogram bin and τ=card($\mathcal{R}$) is the number of pixels in the region $\mathcal{R}$. $p_{i|j}$is the conditional probability of the selected *j*th pixel belonging to the *i*th color bin. It is expressed as follows:

$$p_{i|j} = \begin{cases} 1, if \ the \ jth \ pixel \ is \ quantized \ into \ the \ ith \ color \ bin \\ 0, otherwise \end{cases} \tag{3}$$

2.3 Region classification

In this stage of our method, we dispose of segmented regions $\mathcal{M}_{\text{SRM}}$ obtained via the SRM algorithm and which are characterized using the RGB color histogram as local image region descriptor. the challenge herein is to classify these characterized regions into sky and non-sky regions. To this end, we use the supervised region classification method introduced in [4]. Since this algorithm is a supervised procedure, we have first created two learning databases $\mathcal{B}_{\text{obj}}$and $\mathcal{B}_{\text{back}}$ that are constructed respectively with m distinctive textures of sky regions and n distinctive textures of non-sky regions such as building, road, tree, etc. Figure 3 illustrates the two constructed learning databases $\mathcal{B}_{\text{obj}}$and $\mathcal{B}_{\text{back}}$used in this work. Algorithm

1 consists first in calculating the similarity between the characterized regions ($\mathcal{R}\in\mathcal{M}_{\text{SRM}}$) and those of the two learning databases $\mathcal{B}_{\text{obj}}$and $\mathcal{B}_{\text{back}}$through Hellinger kernel-based distance which represents the cosine of angle between the unit vectors representing the two regions to be compared:

$$(\sqrt{\mathbf{h}_{\mathcal{R}}^{1}},\ldots\ldots,\sqrt{\mathbf{h}_{\mathcal{R}}^{Z}})^{T}\text{and}(\sqrt{\mathbf{h}_{\mathcal{Q}}^{1}},\ldots\ldots,\sqrt{\mathbf{h}_{\mathcal{Q}}^{Z}})^{T}$$

where $\mathbf{h}_{\mathcal{R}}^{i}$is the normalized histogram of the region $\mathcal{R}$, the superscript *i* represents its *i*th element. z = l×l×l = 4096 represents the feature space. Note that the higher the Hellinger distance $\varrho(\mathcal{R},\mathcal{Q})$ (cf. Eq. 4) between regions $\mathcal{R}$ and $\mathcal{Q}$ is, the higher the similarity between them is. That is to say that the angle between the two histogram vectors is very small involving that their histograms are very similar.

$$\varrho(\mathcal{R},\mathcal{Q})=\sum_{i=1}^{z}\sqrt{\mathbf{h}_{\mathcal{R}}^{i}.\mathbf{h}_{\mathcal{Q}}^{i}} \tag{4}$$

After that, the algorithm assigns an unknown region $\mathcal{R}$ to a class $\mathcal{C}_{\text{n}}$, if the average of the *K* first high similarity measures calculated between the region $\mathcal{R}$ and the regions of the learning database corresponding to the class $\mathcal{C}_{\text{n}}$is maximal, i.e,

Algorithm 1 Region classification algorithm.

Require: I ← The set $M_{\text{SRM}}$of segmented regions.
1: $\mathcal{B}_{\text{obj}}$← learning database of sky regions.
2:$\mathcal{B}_{\text{back}}$← learning database of non-sky regions (building, road, tree, etc.).
3:
4: Calculate the local image region descriptor for all regions of $M_{\text{SRM}}$and for those composing $\mathcal{B}_{\text{obj}}$and $\mathcal{B}_{\text{back}}$.
5: **for** each candidate region $\mathcal{R}\in M_{\text{SRM}}$do
6: Calculate $\Lambda_{obj}^{R}=\{\varrho(\mathcal{R},\mathcal{Q}_i);(\mathcal{Q}_i)_{i=1..m}\in\mathcal{B}_{\text{obj}}\}$, the similarity vector between $\mathcal{R}$ and $\mathcal{B}_{\text{obj}}$. $\varrho(\mathcal{R},\mathcal{Q}i)$ is the similarity between R and the region $\mathcal{Q}_i\in\mathcal{B}_{\text{obj}}$.
7: Calculate $\Lambda_{back}^{R}=\{\varrho(\mathcal{R},\mathcal{Q}_j);(\mathcal{Q}_j)_{j=1..n}\in\mathcal{B}_{\text{back}}\}$, , the similarity vector between $\mathcal{R}$ and $\mathcal{B}_{\text{back}}$. $\varrho(\mathcal{R},\mathcal{Q}j)$ is the similarity between R and the region $\mathcal{Q}_j\in\mathcal{B}_{\text{back}}$.
8: Get the order of Λ_{obj}^{R} and Λ_{back}^{R} by decreasing sorting;
9: Calculate $\varphi_{obj}^{\mathcal{R}}=\frac{\sum_{i=1}^{K}\varrho(\mathcal{R},\mathcal{Q}_i)}{K}$, K ≤ m, the mean of the K first elements of $\Lambda_{\text{obj}}^{\mathcal{R}}$.
10: Calculate $\varphi_{back}^{\mathcal{R}}=\frac{\sum_{i=1}^{K}\varrho(\mathcal{R},\mathcal{Q}_j)}{K}$ K , K ≤ n, the mean of the K first elements of $\Lambda_{\text{back}}^{\mathcal{R}}$.
11: **if** ($\varphi_{\text{obj}}^{\mathcal{R}}\geq\varphi_{\text{back}}^{\mathcal{R}}$) **then**
12: Thee region $\mathcal{R}$ has the maximal similarity with $\mathcal{B}_{\text{obj}}$, it is then classified as a part of sky.
13: **else**
14: The region $\mathcal{R}$ has the maximal similarity with $\mathcal{B}_{\text{back}}$, it is then classified as a part of background.
15: **end if**
16: **end for**
return : The Final segmentation map.

$$C^{*}(\mathcal{R})=\underset{C_n\in C}{\text{argmax}}\frac{1}{K}\sum_{i=1}^{K}\varrho(\mathcal{R},\mathcal{Q}_i),\mathcal{Q}_i\in B_n,n=1,2 \tag{5}$$

where $\mathcal{B}_1$and $\mathcal{B}_2$are the learning databases corresponding to the classes $\mathcal{C}_1$and $\mathcal{C}_2$, $\mathcal{R}$ is a query, and ϱ is the similarity measure.

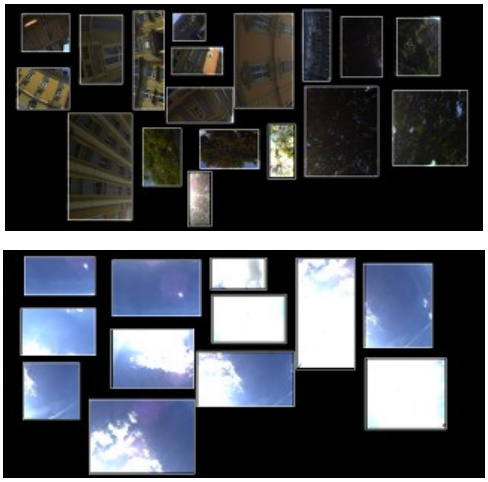


Figure 3: Example of learning databases used in this work. From top to bottom: learning database $\mathcal{B}_{back}$of background (vegetation, building, etc) and learning database $\mathcal{B}_{obj}$of sky.

3 Segment Analysis Technique

The region classification method introduced previously allows achieving good classification results. Unfortunately, in some challenging cases like the presence of buildings with the same color as sky region (generally blue and white colors), windows with mirrors or even mirrored buildings that reflect the color of sky region, the method performs worse. Indeed, the method is failing to distinguish between building and sky regions with the same color using only color information in the region descriptor, it is then accompanied with false positives. Figure 4 illustrates some false detections generated by Algorithm 1 described above. To deal with this shortcoming and therefore improve sky region detection results, we propose to couple Algorithm 1 with a segment analysis procedure. This strategy is based on a detection of segments on which we establish two hypotheses. The first one supposes that the confusion between the vegetation and sky regions is highly impossible via Algorithm 1. the second hypothesis postulates that buildings, referring to their constitution, contain a significant number of segments. The final sky detection strategy we have developed consists of four steps described below:

(1) Detection of segments in the image.
(2) Creation of an optimized mesh.
(3) Conversion into a density map.

Region classification via the whole proposed approach

Figure 4: False detections generated by Algorithm 1.

3.1 Segments detection

This step constitutes the cornerstone of the proposed alternative to overcome the shortcoming of Algorithm 1 and thus will improve the system reliability in terms of classification of the content of fisheye images into two classes (sky, not sky). Many methods of segments detection exist in the literature like Hough transform, Dseg, LSD [13], etc., however many of them make their parameters optimized for a targeted application. According to one of our application constraints, related to the optical sensor used: a color camera with a fisheye lens providing a field of view of more than 180 degrees with significant distortions, we have opted to use the algorithm LSD which others the best results. The LSD (Line Segment Detector) method is a segment detection technique that focuses on the orientation of the gradient and a strict selection of the segments detected. Figure 5 illustrates an example of segments detection on both grayscale and geodisk reconstruction by dilation (GRD) images. Note that the detection results obtained with the LSD algorithm in the framework of our application are promising since we have a large number of segments detected in the building regions and little in the other regions (vegetation, trees, etc.) of the image.

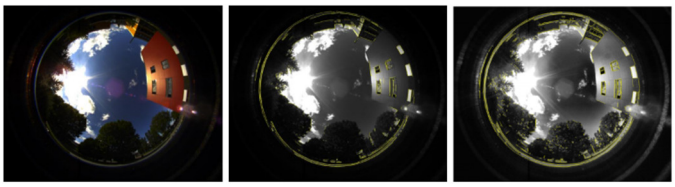


Figure 5: Detection of segments with LSD. From left to right: the original image, detected segments on grayscale and on GRD images.

3.2 Creation of an optimized mesh

As pointed out in the previous step, we have obtained a high density of segments detected in the buildings but also on their contours. The next step consists in connecting different segments to create shapes with closed contours which will be labeled to create a map of regions coinciding with building areas. Œe objective herein is to refine the classification results obtained by Algorithm 1, i.e., reclassify building regions which are classified as true positive (sky region). For that, the creation of the mesh concerns only the detected segments composed of non-sky pixels. That being said, the segments having one of their extremities present in sky area are removed to avoid unnecessary connections of segments in sky regions. During the segment connection step, only the segments where the corresponding extremities are very close to each other in a Euclidean distance sense are connected by a new segment. The tests we have carried out conclude that a value of 20 pixels (distance between extremities of two segments candidates to connection) gives good results. Figure 6 illustrates an example of mesh obtained from segments detected in an image. As can be seen from this figure, the mesh result shows a large concentration of small-sized regions in the building areas.

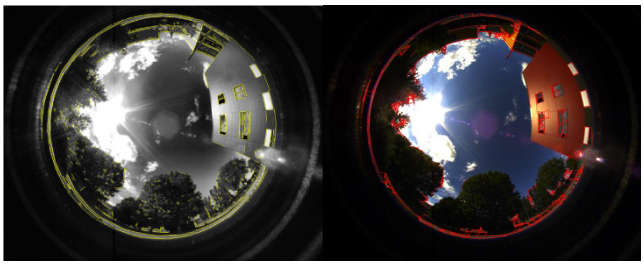


Figure 6: Illustration of segment detection (left), and creation of a mesh (right).

3.3 Conversion into a density map

The so called density map is obtained by filling all the holes in the mesh obtained in the previous step. Figure 7 illustrates the density map corresponding to the mesh of Figure 6. These results, which confirm the hypotheses previously mentioned, are therefore very satisfactory.

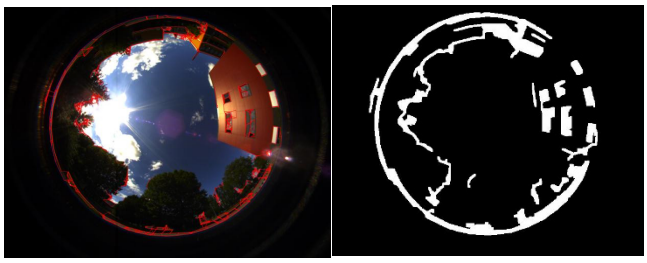


Figure 7: Illustration of the creation of a mesh (left), density map (right).

3.4 Region classification via the whole proposed approach

It is a question herein to define the class to which each region $\mathcal{R}$ belongs among the two possible classes: sky and non-sky, from the region map and the density map previously calculated (cf. Figure 8). The use of the density map allows each region to be assigned a density percentage, i.e., the ratio between the number of pixels belonging to the density map within the candidate region $\mathcal{R}$ and the total number of pixels of the same region. In order to optimize and strengthen decision-making on classes to which image regions belong, we start by considering that each region $\mathcal{R}$, initially classified as sky region via Algorithm 1, whose density percentage is less than 90% of the image, will be automatically classified as sky region. In contrary, if this percentage is greater than 90%, whatever the classification result of Algorithm 1 (sky or non-sky), the region will be classified as non-sky region. Let $\mathcal{R}\in$ region map and P_R is the density percentage. The class to which the region $\mathcal{R}$ belongs is determined using Eq 6. Algorithm 2 illustrates the steps of the whole proposed approach.

$$R = \begin{cases} sky & if \quad P_R < 90\% \; et \; \Omega(R) = R_{sky} \\ non - Sky & if \quad P_R >= 90\% \end{cases} \tag{6}$$

where Ω is the set of obtained sky/non-sky regions via Algorithm 1.

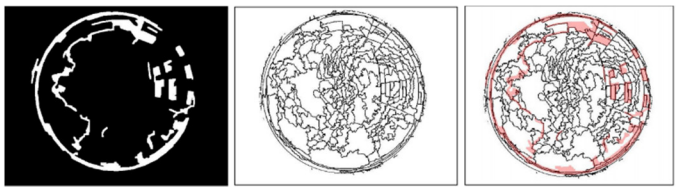


Figure 8: Merge of density and region (cf. Figure 2) maps. From left to right: density map, region map and the obtained merging result.

Algorithm 2 Proposed approach

Require: I ← input image.

M_{SRM} ← SRM Area Map.

Ω ← obtained sky/non-sky regions via Algorithm 1.

1: I_{GRD} ← Apply the GRD contrast to the gray level (I).

2: Detection of segments in I_{GRD}image with multiple resizing.

3: Creation of an optimized mesh.

4: CD ← Conversion of the mesh to a density map.

5: **for** each candidate region $\mathcal{R}\in M_{SRM}$**do**

6:Calculate the density percentage $P_{\mathcal{R}}$ = Number of CD pixels/number of pixels in $\mathcal{R}$

7: **if**P_R< 90% et Ω(R) = R_{sky}**then**

8: The region $\mathcal{R}$ is classified as a sky region.

9:**else**

10: The region $\mathcal{R}$ is classified as a non-sky region.

11: **end if**

12: **end for**

13: The classification process is complete.

14: **return**The final map of detection

4 Experimental Results

In this section, we are interesting in assessing the ability of the proposed approach to deal with the detection of sky zones from fisheye images. We begin by a qualitative evaluation of the proposed method using representative test images. Figure 9 illustrates an example of visual comparison of the region classification results. From top to bottom: classified image into two classes (sky and non-sky) obtained by Algorithm 1 and classification result obtained by the proposed approach. In Figure 9, the sky boundaries are drawn with green color and superimposed upon the initial images. Basing on visual evaluation of these results, one can see clearly that, due to radiometric similarity between some pixels belonging to building and sky regions, several non-sky parts are classified as true positive (sky region). In contrary, from the second row of Figure 9, it is clear that the proposed approach demonstrates excellent accuracy in terms of sky boundary extraction. This means that the majority of the sky zones present in the images are detected with good boundary delineation. Indeed, our method shows reliable results across complex environment composed of sky, building, road areas, vegetation, etc.

As for quantitative evaluation, we use measures widely employed in evaluating effectiveness. They constitute a useful and accepted tool in object recognition field [12]. The 150 fisheye images of the tested dataset were first manually delineated. then, they are used as a ground truth image segmentation to assess the accuracy of automated sky regions extraction. The obtained results and reference ones are compared pixel-by-pixel. Each pixel in the images is categorized as one of the four possible outcomes:

(1) True positive (TP): Both manual and automated methods label the pixel belonging to sky.

(2) True negative (TN): Both manual and automated methods label the pixel belonging to background.

(3) False positive (FP): The automated method incorrectly labels the pixel as belonging to sky.

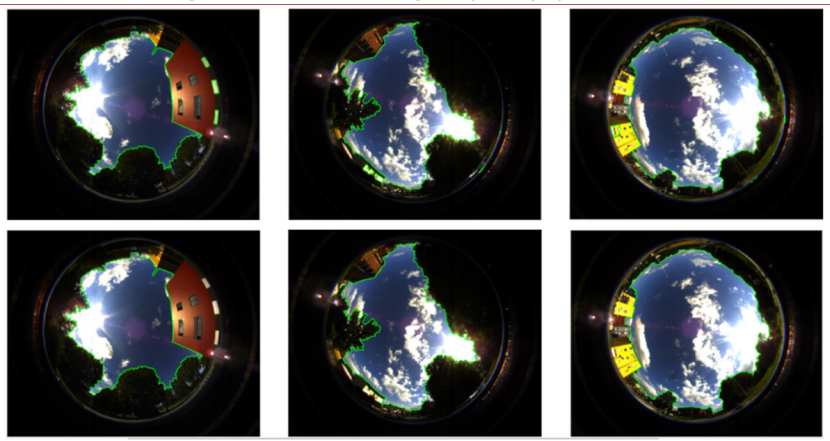


Figure 9: Visual comparison of region classification results. From top to bottom: classified image into two classes (sky and non-sky) obtained by Algorithm 1 and classification result obtained by the proposed approach.

Table 1: Evaluation results using Algorithm 1 and the proposed approach (Algorithm 2)

Method	Recall (%)	Precision (%)	F-measure (%)	Accuracy (%)	MMC (%)
Algorithm 1	98.66	96.36	97.48	99.00	0.97
Proposed approach (Algorithm 2)	98.90	97.48	98.18	99.10	0.98

(4) False negative (FN): The automated method does not correctly label the pixel truly belonging to sky.

To examine detection performance, the number of pixels that fall into each of the four categories TP, TN, FP, FN are determined, and the following measures are computed:

$$Recall = \frac{TP}{P} \times 100 \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{8}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{TP + FP} \tag{9}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP} \times 100 \tag{10}$$

$$Recall = \frac{(\text{T P} \times \text{T N}) - (\text{FP} \times \text{FN})}{\sqrt{(\text{T P} + \text{FP}) \times (\text{T P} + \text{FN}) \times (\text{T N} + \text{FP}) \times (\text{T N} + \text{FN})}} \tag{11}$$

Table 1 summarizes the average (on the dataset) of the performance indicators recall, precision, F1 measure, Accuracy and MCC according to Algorithm 1 used alone on the one hand and with the proposed segment analysis technique. It can be clearly seen that the proposed method maximizes the performance indicators compared to Algorithm 1 used alone. Indeed, the proposed approach demonstrates its effectiveness as it shows superior performances, which confirms hence the conclusion given from the analysis of Figure 9.

5 Conclusion

In this paper, an automatic sky region detection method is proposed. The procedure we proposed includes a region based classification method and a segment analysis based technique. The first method is used in order to classify segmented regions via Statistical Region Merging, characterized by RGB color descriptor, into two classes (sky and non-sky regions). Œe second method is introduced in order to refine the classification results, especially the regions belonging to building area which were mistakenly classified by the first method as true positive (sky region). As highlighted by the obtained results, the proposed approach is therefore able to distinguish perfectly the sky/non-sky areas in a fisheye image. For future works, we plan to improve the effectiveness of the proposed approach by proposing new color descriptors and evaluate the impact of other segment detection methods on the quality of the classification results.

REFERENCES

[1] Dhouha Attia. 2014. Segmentation d'images par combinaison adaptative couleur-texture et classification de pixels. : Applications la caractérisation de l'environnement de réception de signaux GNSS. PhD thesis, Université de ´ technologie Belfort-Montbéliard(Jun 2014).

[2] Y El merabet, Y Ruichek, S Ghaffarian, Z Samir, T Boujiha, R Touahni, and R. Messoussi. 2016. Horizon Line Detection from Fisheye Images Using Color Local Image Region Descriptors and Bhattacharyya Coefficient-Based Distance. In International Conference on Advanced Concepts for Intelligent Vision Systems. Springer International Publishing. 58–70.

[3] Juliette Marais et Cyril Meurie. 2013. Quantify et improve GNSS quality of service in land transportation by using image processing. First CNES-ONERA Workshop on Earth-Space Propagation, hal-00912681 (December 2013).

[4] A. El Idrissi, Y. El Merabet, Y. Ruichek, R. Touahni, A. Sbihi, C. Meurie, and A. Moussa. 2015. A Multiple-Objects Recognition Method Based on Region Similarity Measures: Application to Roof Extraction from Orthophotoplans. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 11 (Novembre 2015).

[5] J. Marais, C. Meurie, S. Ambellouis, and Y. Ruichek. 2013. Projet CAPLOC. Combinaison de l'analyse d'images et la connaissance de la propagation des signaux pour la Localisation. (Agust 2013).

[6] L. Matthies, M. Maimone, A. Johnson, C. Yang, R. Willson, C. Villalpando, S. Goldberg, A. Huertas, A. Stein, and A. Angelova. 2007. Computer vision on mars. Int. J. Comput. Vis. 75, 1 (2007), 67-92.

[7] J. Ning, L. Zhang, D. Zhang, and C. Wu. 2009. Interactive image segmentation by maximal similarity based region merging. (Mars 2009).

[8] R. Nock and F Nielsen. 2004. Statistical region merging. IEEE Trans. Pattern Anal. Machine Intelligence 26, 11 (2004), 1452–1458.

[9] S. Ramalingam, S. Bouaziz, P. Sturm, and M. Brand. 2010. SKYLINE2GPS: Localization in Urban Canyons Using Omni-Skylines. IEEE International Conference on Intelligent Robots and Systems (2010), 3816–3823.

[10] Z. B Sbeh, L. D Cohen, G Mimoun, and G Coscas. 2001. A New Approach of Geodesic Reconstruction for Drusen Segmentation in Eye Fundus Images. IEEE Transactions on medical imaging 20, 12 (Decembre 2001), 1321–1333.

[11] Y. Shang, G. Li, Z. Luan, X. Zhou, and G Guo. 2016. Sky detection by effective context inference. Neurocomputing 208 (2016), 238–248.

[12] J. A Shufelt. 1999. Performance evaluation and analysis of monocular building extraction from aerial imagery. IEEE Trans. Pattern Anal. Machine Intell 21 (1999), 311–326.

[13] R. G Von Gioi, J Jakubowicz, J. M Morel, and G Randall. 2012. LSD: a line segment detector. Image Processing On Line 2 (Mars 2012), 35–55.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Robust Fuzzy Neural Network Sliding Mode Control for Wind Turbine with a Permanent Magnet Synchronous Generator

K. Belamfedel Alaoui, S.Sefriti and I.Boumhidi
LESSI Laboratory, Department Of Physics, Faculty Of Sciences, Sidi Mohammed Ben Abdellah University, Morocco
khawala.belamfedelalaouikhaoula@usmba.ac.ma; selma.sefriti@usmba.ac.ma; iboumhidi@hotmail.com

ABSTRACT

In the present paper, we are interested in the contribution of wind power to the electricity supply in power systems of small sized isolated communities. A robust fuzzy neural sliding control (FNSC) is proposed to track the maximum power point for an isolated wind energy conversion system using a permanent magnet synchronous generator (PMSG) with a hysteresis rectifier connected to a DC load. The turbine is controlled by a sliding mode controller (SMC) to reach the maximum power level. The main objective of the control is to adjust the rectifier voltage to provide it to the DC load as well as to maintain the maximum power extraction. In presence of large uncertainties and wind speed variations, the traditional SMC produces the chattering phenomenon due to the higher needed gain. In order to reduce this gain, FNSC is used for the estimation of the unknown part, thus provide lower gain. The stability of the proposed FNSC is analyzed by LYAPUNOV theory, simulations results are presented and the proposed control performance is shown by the comparison with the conventional SMC.

Keywords—variable speed wind turbine; permanent magnet synchronous generator; sliding mode control; fuzzy neural sliding mode control

1 Introduction

Wind energy is one of the most important renewable sources which are used in power systems either for stand alone or for grid connected applications. The use of wind energy to generate electricity is now accepted with a large industry manufacturing and installing thousands of Wind Turbines (WT) each year. Although, there are exciting new developments of wind turbines to make them more profitable [1]. Practically, there are two main types of vertical axis wind turbines: fixed speed and variable speed [2]. In this study, we consider the case of variable speed, due to its great ability in the extraction of energy [3].This work is devoted to the electrical part with the aim of designing a robust controller to adjust the rectifier voltage, to keep at the same time the captured energy at its maximum and to reduce the chattering phenomenon due to the higher needed gain. . The output tracking of non-linear WT system with uncertainties using conventional Proportional Integral Derivative (PID) controller is quite complex, since conventional PID controllers are tuned to a particular operating point. This degrades the performance of PID controller under non-linearities. Developing a simple and robust controller for uncertain non-linear WT system is important to overcome this problem.

DOI: 10.14738/tmlai.54.3336
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3336

Sliding mode control ([4]-[6]), emerges as an especially suitable option to deal with variable speed wind turbine. This is due to its robustness with uncertainties and disturbances. Therefore, when the uncertainties are large, a high discontinuous control gain is needed and higher amplitude of chattering is produced.

To control nonlinear uncertain dynamic systems, adaptive fuzzy neural control has been used to solve approximation functions problems. The fuzzy neural network (FNN) has proved in many applications that it possesses the propriety of universal approximation [7].

The designed method is a combination of conventional SMC and FNN with online adaptation of the parameters to adjust the stator voltages of the PMSG. The proposed control combines the advantages of the FNN and those of SMC because of its ability to predict unknown part in the equivalent control term, so the gain used in the robust control term becomes smaller which helps the network error to be compensated.

This study is organized as follows. The next section presents the system modeling and control objectives. Section 3 shows the design of the proposed control strategies for power capture optimization. In Section 4, simulation results are provided to demonstrate the robust control performance of the proposed approach. Finally, a concluding remark is given in section 5.

2 System Modeling and Control Objectives

The proposed system contains a variable speed wind turbine coupled directly to a PMSG, a rectifier and a DC load as it is illustrated in figure 1.

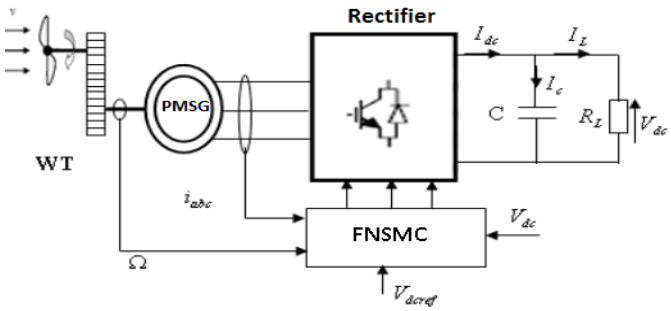


Figure 1: The structure of the studied system

2.1 Wind Turbine Modeling

Wind energy across a surface S depends on the cube of the wind speed v and the density of the air ρ .This energy is given by:

$$P_v = \frac{1}{2} S \rho v^3 \tag{1}$$

We are interested in a vertical axis turbine where the active surface is in function of the geometric dimensions of the wing shown in figure 2:

$$S = 2.R.H \tag{2}$$

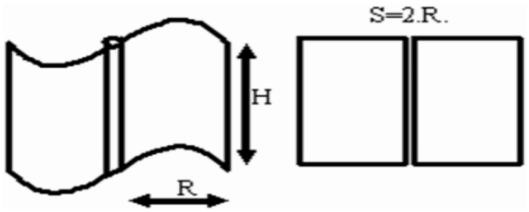


Figure 2: Geometric dimensioning Savonius wing

R is the radius of the rotor and $H$is the height of the turbine

The aerodynamic power captured by the rotor is given as follows:

$$P_a = \frac{1}{2}\rho\pi R^2 C_p(\lambda,\beta)v^3 \tag{3}$$

The power coefficient C_p depends on the blade pitch angle β and the tip-speed ratio λ which is defined as follows:

$$\lambda = \frac{\omega_t R}{v} \tag{4}$$

With ω_t is the rotor speed.

The power coefficient $C_p(\lambda,\beta)$ is a nonlinear function of λ and β [8]:

$$C_p(\lambda,\beta) = C_1\left(\frac{C_2}{\lambda_i} - C_3\beta - C_4\right)\exp\left(-\frac{C_5}{\lambda_i}\right) + C_6\lambda \tag{5}$$

Where:

$$\frac{1}{\lambda_i} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1} \tag{6}$$

And the parameters c_i are known constants without units.

In this paper, we focus on the modeling of the electrical part of the variable speed wind turbine, which is presented by PMSG.

The equations of PMSG in a d-q rotating frame linked to the rotor as follows [9]:

$$\begin{cases} V_{sd} = R_s I_{sd} + L_{sd}\frac{d}{dt}I_{sd} - pwL_{sq}I_{sq} \\ V_{sq} = R_s I_{sq} + L_{sq}\frac{d}{dt}I_{sq} + pwL_{sd}I_{sd} + pw\,\phi_f \end{cases} \tag{7}$$

With:

R_s : is the resistance of the stator windings,

I_{sd} I_{sq} : are the stator currents in the Park rotating frame, V_{sd} V_{sq} : are the stator voltage in the Park rotating frame, L_{sd} L_{sd} : are Inductances along the direct and the quadrature axes ,

p : is the number of pole pairs,

$\omega = p.\Omega$: Voltage pulsation (rad/s),

And ϕ_f is the magnetic flux created by the PMSG through the stator windings.

The electromagnetic torque in the rotating frame is given by:

$$T_e = \frac{3}{2} p \left[\left(L_{sd} - L_{sq} \right) I_{sd} I_{sq} - \phi_f I_{sq} \right] \quad (8)$$

2.2 Control Objectives

Below the nominal power, the main control objective is to define a resistive electromagnetic torque T_e to be provided by the generator in order to optimize the energy captured from the wind. The power coefficient curve $C_p(\lambda, \beta)$ has a unique maximum which corresponds to the optimal wind energy (see Figure 3):

$$C_p(\lambda_{opt}, \beta_{opt}) = C_{popt} \quad (9)$$

The rotor, thus, provides maximum aerodynamic power only to the tip-speed λ_{opt} :

$$\lambda_{opt} = \frac{\omega_{topt} R}{v} \quad (10)$$

To maximize the captured energy of the wind, the variables λ and β must be maintained at their optimal values in order to ensure maximum value of C_p . So, the blade pitch angle is fixed at its optimal value β_{opt} . The tip-speed λ depends on both of the wind speed v and the rotor speed ω_t . As the wind speed is not a controllable input, the rotor speed ω_t must be adjusted by T_e , to track the optimal reference given by:

$$\omega_{topt} = \frac{\lambda_{opt}}{R} v \quad (11)$$

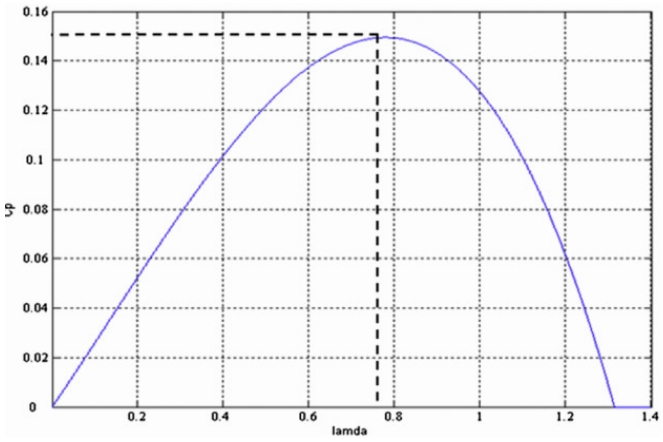


Figure 3: Power versus tip speed ratio for $\beta = \beta_{opt}$

3 Fuzzy Neural Network Sliding Mode Control strategy

3.1.1 Conventional Sliding Mode Controller (SMC)

SMC is one of the most important robust control approaches with respect to system dynamics and invariant to uncertainties. LYAPUNOV stability approach is used in SMC to keep the nonlinear system under control. Our target is to control the stator current of the PMSG and the DC voltage even with the wind speed variations.

Let define the control vector $u = [V_{sd} V_{sq}] = [u_1 u_2]$ and the output vector $y = [V_{dc} I_d]$ and the system state $x = [I_{sd} \ I_{sq}]^T$

The model described by (7) can be rewritten in the state space as follows:

$$\begin{cases} \dot{X} = f(x) + G(x)U + \xi \\ y = [y_1 y_2] \end{cases} \tag{12}$$

$$\dot{X} = \begin{bmatrix} a_{11} x_1 + a_{12} x_2 \\ a_{21} x_1 + a_{22} x_2 + a_{23} \end{bmatrix} + \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix} U + \xi$$

With

$$a_{11} = \frac{R_s}{L_{sd}}, \ a_{12} = \omega \frac{L_{sq}}{L_{sd}}, \ a_{21} = \frac{R_s}{L_{sq}}, \ a_{22} = \omega \frac{L_{sd}}{L_{sq}}$$

$$a_{23} = \omega \frac{\phi_f}{L_{sq}}, g_1 = \frac{1}{L_{sd}}, g_2 = \frac{1}{L_{sq}} \text{ and } \xi \text{ the unknown term.}$$

In the following, we have two tracking errors since we have two control laws. The tracking errors of the system are defined as:

$$e_1(t) = I_{sd}(t) - I_{sd\_ref}(t) \tag{13}$$

$$e_2(t) = I_{sq}(t) - I_{sq\_ref}(t) \tag{14}$$

To design the SMC there are two steps. First finding the sliding surface second, developing the control law.

Generally, the sliding surface is defined as:

$$\sigma(t) = \left(\gamma + \frac{d}{dt}\right)^{r-1} e(t) \tag{15}$$

where r is the order of the system and γ is a positive constant.

The relative degree of the system (12) is 1. Then, the sliding surfaces can be defined as:

$$\sigma_1 = e_1 \tag{16}$$

$$\sigma_2 = e_2 \tag{17}$$

Differentiating σ with respect to time, we have:

$$\dot{\sigma}_1 = f_1(x) + g_1 u_1 - \dot{I}_{sd\_ref} + \xi \tag{18}$$

$$\dot{\sigma}_2 = f_2(x) + g_2 u_2 - \dot{I}_{sq\_ref} + \xi \tag{19}$$

To guarantee the existence of sliding mode on a switching surface in finite time, the condition given below has to be satisfied [5]:

$$\sigma \dot{\sigma} < 0 \tag{20}$$

The control law that satisfies Eq. (20) is given by [9]:

$$u_1 = -\frac{f_1(x) - \dot{I}_{sd\_ref}}{g_1} - \frac{k_1}{g_1} sign(\sigma_2) \tag{21}$$

$$u_2 = -\frac{f_2(x) - \dot{I}_{sq\_ref}}{g_2} - \frac{k_2}{g_2} sign(\sigma_2) \tag{22}$$

Where k_1 and k_2 are the positive switching gains to compensate the uncertainties. With $sign(.)$ is the sign function, given by:

$$sign(\sigma) = \begin{cases} 1 & if\ \sigma > 0 \\ 0 & if\ \sigma = 0 \\ -1 & if\ \sigma < 0 \end{cases} \tag{23}$$

To eliminate the chattering effect caused by the discontinuous control law, the boundary layer approach can be used. The control becomes as follows:

$$u_{1,2} = -\frac{f_{1,2}(x) - \dot{I}_{d,q\_ref}}{g_{1,2}} - \frac{k_{1,2}}{g_{1,2}} sat(\sigma_{1,2}) \tag{24}$$

Where $sat(.)$ is the saturation function, given by:

$$sat(\sigma) = \begin{cases} \sigma / L & if\ |\sigma| < L \\ sign(\sigma) & \text{otherwise} \end{cases} \tag{25}$$

With L is the boundary layer thickness.

3.2 Fuzzy Neural Network Representation

The fuzzy neural network considered is formed by five layers. The first one is for the inputs; the second one is formed by the memberships, the third one contains the rules; the fourth one is the normalized and the output layer receiving the weighted outputs. The weights are adjustable to estimate the unknown term ξ.

This network is illustrated in figure 4.

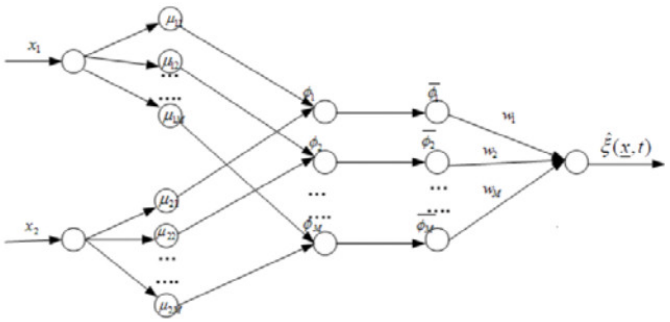


Figure 4 The proposed structure of the fuzzy neural network

We describe in the following the function of each layer:

Layer1: it's the input layer which contains the input variables $x_{1,2}$, transmits directly input values to the next layer.

Layer2: in this layer, each node performs a membership function and acts as a unit of memory. The Gaussian function is adopted as the membership function. For the k[th] input, the corresponding net input and output of the l[th] node can be expressed as:

$$\mu_k^l(x_k) = \exp\left(-\frac{(x_k - c_k^l)^2}{(\sigma_k^l)^2}\right) \tag{26}$$

Where c_k^l and σ_k^l; $l = 1, 2, \ldots, M$ are the mean and standard deviation of the Gaussian function of the l[th] partition for the k[th] input variable x_k, respectively, and M is the total number of fuzzy rules.

Layer3: this layer represents one fuzzy logic rule and performs prediction matching of a rule. The output of a rule node in this layer is calculated by the product operation as follows:

$$\phi(x) = \prod_{k=1}^{2} \mu_k^l(x_k) \tag{27}$$

Layer4: in this layer, each of these firing strengths of the rules is compared with the sum of all the firing strengths. Therefore, the normalized firing strengths are computed in this layer as,

$$\bar{\phi}(x) = \frac{\phi(x)}{\sum_{l=1}^{M} \phi(x)} \tag{28}$$

Layer5: this is the output layer calculates the summation of its input values from the previous.

$$\hat{\xi}(x) = \sum_{l=1}^{M} w_l \bar{\phi}(x) \tag{29}$$

The FNN prediction error is w given in the next section.

The aim of the learning is to adjust the weights of, w_{ς}^{l} , σ_{ς}^{l} and c_{ς}^{l} . The online Learning algorithm is a gradient descent (GD) search algorithm in the space of the network parameters. The essence of (GD) consists of iteratively adjusting the weights the direction opposite to the gradient of error, so as to reduce the discrepancy according to:

$$\dot{w}_{l} = \eta_{wl} w \phi \tag{30}$$

Where η_{wl} is the learning rate for w_{y}^{l} .

The updated laws of c_{ς}^{l} and σ_{ς}^{l} also can be obtained by gradient descent search algorithm:

$$\dot{c}_{k}^{l} = 2\eta_{c_{k}^{l}} w \phi w_{l} \frac{(x_{k} - c_{k}^{l})}{(\sigma_{k}^{l})^{2}} \tag{31}$$

And

$$\dot{\sigma}_{k}^{l} = 2\eta_{\sigma_{k}^{l}} w \phi w_{l} \frac{(x_{k} - c_{k}^{l})^{2}}{(\sigma_{k}^{l})^{3}} \tag{32}$$

Where $\eta_{c_k^l}$ and $\eta_{\sigma_k^l}$ are the learning-rate parameters of the mean and the standard deviation of the Gaussian functions.

3.3 Proposed Fuzzy Neural Sliding Controller (FNSC)

We are interested in implementation of a robust control of nonlinear system, uncertain and disrupted as (12).

To develop the control law, we propose to approximate the unknown functions ξ by the FNN presented in figure 4.

Let denote the approximation error as: $w = \xi - \hat{\xi}$

With $|w| < w^{*}$ and w^{*} is the upper bound of the network error supposed known.

Theorem: Consider the system described by (12) in the presence of uncertainties. If the system control is designed as:

$$U_{1,2} = g_{1,2}^{-1}\left(- f_{1,2}(x) - \hat{\xi} + \dot{I}_{d,q-ref} - Ksat(\sigma_{1,2}) \right)$$

Where $w^{*} + \eta < k$; with k is a small positive vector values

The trajectory tracking error will converge to zero in finite time.

Proof: we consider the LYAPUNOV function:

$$V = \frac{1}{2}\sigma^{2} \quad \text{Then} \quad \dot{V} = \sigma\dot{\sigma}$$

$$\dot{V} = \sigma\left(f_{1,2}(x) + g_{1,2}(x)u_{1,2} - \dot{I}_{d,q-ref} + \xi - Ksat(\sigma) \right)$$

By replacing the expression of u given in the theorem we have:

$$\dot{V} = \sigma\left(\hat{\xi} - \xi - ksat(\sigma) \right)$$
$$= \sigma\left(w - ksat(\sigma)\right)$$
$$= \sigma w - k\,\sigma\, sat(\sigma)$$
$$\leq |\sigma|\,\|w\| - k\,\sigma\, sat(\sigma) < |\sigma|\,w - k\,\sigma\, sat(\sigma)$$

By choosing $w^* + \eta < k$ we have:

For any $L > 0$, if $|\sigma| > L$, $sat(\sigma) = sign(\sigma)$ the function $\dot{V} = (w^* - k)|\sigma| < 0$. However, in a small *L*-vicinity of the origin [5], (boundary layer), $sat(\sigma) = \frac{\sigma}{L}$ is continuous, the system trajectories are confined to a boundary layer of sliding mode manifold $\sigma = 0$.

4 Simulation Results

In order to make a comparison between the proposed control strategies, different simulations are considered for the case of the PMSG with:

A wind speed profile of 10 m/s means value.

$C_1 = 0.5176,\ C_2 = 116,\ C_3 = 0.4,\ C_4 = 5,\ C_5 = 21,\ C_6 = 0.0068$.

$R = 0.5m,\ \rho = 1.29kg/m^3,\ H = 2m\ S = 2m^2$

$L_s = 0.0263H$, $R_s = 1.4\Omega$ $C_{p\max} = 0.15$

$\phi_f = 0.15Wb$, $R_L = 80\Omega$, $C = 500\mu F$

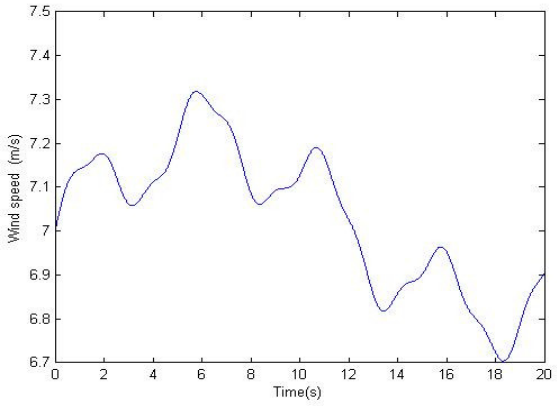


Figure 5 Wind speed profile(m/s)

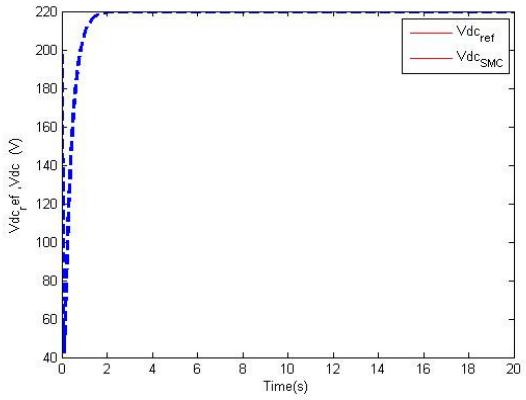


Figure 6 SMC DC voltage and reference voltage without uncertainties

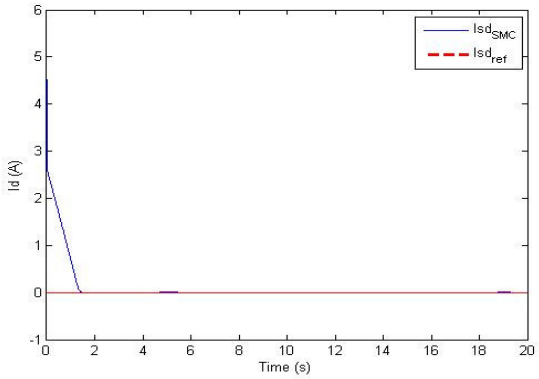


Figure 7 Direct current by traditional sliding mode control without uncertainties

Three membership functions have been used for each of the two inputs (I_{sd} I_{sq}) of the FNN.

Let now consider the uncertainties as a vector random noise with the magnitude equal.

$$\xi = 10 + 0.3 \times \text{rand}()$$

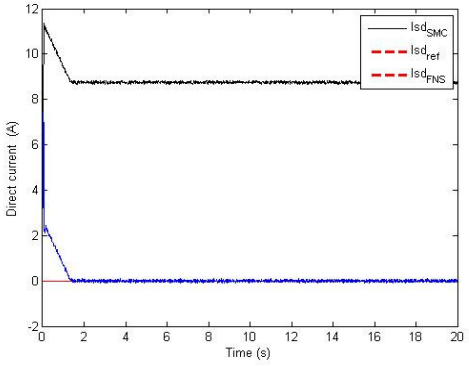


Figure 8 Direct current by the fuzzy neural sliding control and SMC

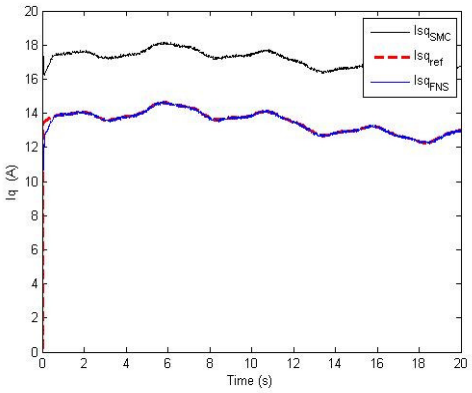


Figure 9 stator current by the fuzzy neural sliding control and SMC

Figure6 presents the response of the DC voltage by the proposed control and the desired one; Figure7 presents the direct current controlled by traditional sliding mode control without any disturbances; Figure8 and Figure9 present a comparison between the traditional sliding mode control and the proposed fuzzy neural sliding mode control in the presence of disturbances. From these figures, it can be seen that the best performance is obtained when the proposed approach is applied.

5 Conclusion

In this study, a control of a permanent magnet synchronous generator and the DC voltage feeding a hysteresis rectifier is presented. The proposed control strategy is based on nonlinear predictive fuzzy neural sliding control. The comparison with the traditional sliding mode control has been realized and

simulation results have shown a good performance of the proposed method to track the optimal reference without any oscillatory behavior. The proposed model has been implemented using Matlab.

REFERENCES

[1] J. F. Manwell, J. G. McGowan and A.L. Rogers, Wind Energy Explained: Theory, Design and Application, New York: John Wiley & Sons,2002.

[2] G. Ofualagba and E.U. Ubeku, "Wind energy conversion system-wind turbine modelling," IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, pp. 1-8 , 2008.

[3] T. Burton, D.Sharpe,N. Jenkins and E. Bossanyi, Wind Energie Handbook. New York: John Wiley & Sons, 2001.

[4] V. I. Utkin, Sliding Modes in Control Optimization, Berlin: Springer-Verlag, 1992.

[5] J. J. Slotine, "Sliding controller design for non-linear systems," International Journal of Control, vol. 40, pp. 421–434 , 1984.

[6] V. Utkin, J. Guldner and J. Shi, Sliding Mode Control in Electromechanical System, London: Taylor & Francis, 1999.

[7] E-M. Boufounas, J. Boumhidi and I. Boumhidi, "Optimal H∞ control without reaching phase for a variable speed wind turbine based on fuzzy neural network and APSO algorithm", International Journal of Modelling, Identification and Control, Vol. 24, No. 2, pp. 100-109, 2015.

[8] H.P. Wang, A. Pintea, N. Christov, P. Borne and D. Popescu , "Modelling and recursive power control of horizontal variable speed wind turbines," Journal of Control Engineering and Applied Informatics, vol. 14, no. 4, pp. 33-41, 2012.

[9] D .Cardenas, "Control of wind turbine using a switched reluance generator". Phd thesis university of Nottingham.1996.

[10] E-M. Boufounas, J. Boumhidi and I. Boumhidi, "Optimal neural network sliding mode control for a variable speed wind turbine based on APSO algorithm", Digests 2nd Annual Conf. IEEE on Complex Systems (WCCS'14) Morocco, p. 419-424, 2014.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Toward Multi-Approach Model for Semi-Automating a Data Warehouse Design from an Ontology

Morad Hajji, Mohammed Qbadou, Khalifa Mansouri
Laboratory: Signals, Distributed Systems and Artificial Intelligence (SSDIA)
ENSET Mohammedia,
University Hassan II of Casablanca, Morocco
morad.hajji@gmail.com, qbmedn7@gmail.com, khmansouri@hotmail.com

ABSTRACT

The proliferation of projects that are part of the semantic Web is truly impressive. In fact, ontologies become increasingly present in information systems, they constitute great data sources that arouse the interest of being analyzed. Ontologies are used for standardizing, structuring and formalizing the Web, Web Service, E-learning systems, and other fields. Regarding multidimensional approaches, researches in this field have focused on direct Data Warehouse conception from an ontology, which do not integrate the intervention of the expert in this process. In this case, the transformations are global and not very customizable; it can reproduce the inherent defects from the data sources into the resulting data warehouse.

In this paper, we propose a new multi-approach model based on the coupling of relational database design approaches from an ontology with Data Warehouse design approaches from a relational database. Our model is semi-automatic allowing Data Warehouse design from an ontology by giving the designer more ability to intervene in this process and closely control the transformations. To assess the usefulness of our approach, we evaluated it by applying it on an example case study. The results of the example show that our approach is more accurate in terms of useful data filtering and adaptation of the multidimensional model to the end-users business-needs.

Keywords-Semantic Web; Business Intelligence; Data Warehouse; Ontology; Database; Coupling; Automatic; Design.

1 Introduction

The primary goal of the Semantic Web is to provide a solution to the structural problem of contemporary Web content. The non-structuring of this content makes its exploitation difficult and inefficient. As advocated by TIM BERNERS-LEE, Semantic Web is intended to promote the formal structuring of Web content and the interoperability between machines themselves and between machines and humans. In fact, the Semantic Web reflects a perception of an environment design that promotes exchanges between software agents and users (humans). The Semantic Web is fundamentally based on the notion of ontology. Ontology reference definition is stated by Gruber [6] and refined by Studer and al. [7]: an ontology is a formal and explicit specification of a shared conceptualization.

DOI: 10.14738/tmlai.54.3337
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3337

Technologies revolving around the Semantic Web have multiplied in a breathtaking way. Indeed, since its polemical birth, the Semantic Web continue to promote the interest of researchers and standards organisms. In particular, W3C (World Wide Web Consortium), which is the source of multiple recommendations such as Ontology Web Language (OWL), Resource Description Framework Schema (RDFS), Resource Description Framework (RDF) and SPARQL (SPARQL Protocol and RDF Query Language). In addition, the notable emergence of numerous systems for the management of ontologies such as KAON [4] and a variety of tools for manipulating ontologies such as the software 'Protégé' [8].

The modern era is experiencing an impressive proliferation of projects that are part of the Semantic Web and thus integrate the notion of ontology. Ontologies are used for the standardization, structuring and formalization of the Web, Web Service, e-learning systems, etc.

Any Semantic Web implementation requires the manipulation of ontologies that are considered as the main pillar of this discipline. In addition, the integration of the Semantic Web into several information systems relating to several domains is growing continually and multiplying day after day. As a result, ontologies have become widely used in many fields and their number and size are extensively increasing. While some ontologies are standardized and normalized, others remain local and specific to the needs of the communities that create them. Ontologies are knowledge bases that store semantically structured information. Thus, ontologies may be interrogated and analyzed in several ways and according to conventional approaches that are based on query languages or non-conventional approaches using multidimensional models.

Concerning multidimensional approaches that deal with the analysis of ontologies, many approaches have been proposedby researchersfor the design of a Data Warehouse from an ontology, as [28], [29], [30] and [31]. However, most of them ignore the end-user business-needs and do not integrate the verification of the results by an expert.Thus, these approaches may not satisfy the specific needs of the end user. On the other hand, they promote the propagation of anomalies from the data source (ontology) to the data warehouse.Therefore, they lead to an inaccurate or inadequate analysis. Moreover, they only deal with ontologies expressed in a specific language such as OWL and RDF. In this case, the transformation is global and not customizable.In fact, it can reproduce the inherent defects in the data sources into the resulting data warehouse. There is a partial satisfaction for the analysis of ontologies. These approaches are not suitable, because loading data from an ontology into a multidimensional scheme can lead to facts related to non-specified or multipledimensions/measures.

The aim of our paper is to propose a multi-approach model for the semi-automatic and controlled design of a Data Warehouse from an ontology. Relational database approaches (RDB) design from an ontology are integrated with approaches of Data Warehouse (DW) design from a relational database. Each integration is a new approach for designing a Data Warehouse from an ontology.

This article is part of our research that aims to develop models in the fields of Semantic Web and Business Intelligence, in particular new models favoring models coupling from these two disciplines. In the course of this work, we have proposed in a previous paper a model that reflects our systemic and integrated perception of the analyzing process of an ontology generated from a documentary corpus [1]. In this article, our main aim is to complete this approach by proposing a model dealing with the ETL (Extraction-Transformation-Loading) and Data Warehousing, which is part of the model proposed in our first article [1].

2 Background

Since the beginning of our research, we have worked on the link between the Semantic Web and Business Intelligence, when our article [1] was developed; our approach has adopted a roadmapto establish a model of a decision-making system adopting ontological sources among the data sources. The new model proposed for this coupling clearly identifies the "Extraction-Transformation-Loading (ETL)" and "Data Warehousing" phase, ensuring the transposition of the data from an ontological source to the multidimensional model.

On the other hand, the approaches of a Data Warehouse design from an ontology differ fundamentally according to the level of representation considered of this ontology. An ontology can be considered as a database consisting of its structure (T-Box) and the data it contains (A-Box). Therefore, the design of the Data Warehouse is based on the T-Box and the loading is based on the A-Box [1].

2.1 Designing a RDB from an ontology

The relational model was initiated by its founder Edgar Frank Codd in 1969. The strength of this model stems from its robust mathematical basis. This model, advocated by large IT organizations as Oracle and Microsoft, is currently present in almost all information systems. At the coupling level between the relational model and the ontologies, we identify mainly at least three orientations:

- Ontology construction from a relational database.
- Relational database construction from an ontology.
- Starting from a relational database and an ontology to design links between their components. It is about the mapping between a database and an ontology.

In the context of our contribution, we are interested in the second type of coupling between an ontology and a database, namely the design of a relational database from an ontology. There are many research studies dealing with this coupling. The work of Gali et al. [11], Vysniauskas and al. [12], Astrova and al. [13], Fankam and al. [14], J. Trinkunas and O. Vasilecas [15], Astrova and al. [16], E. Vysniauskas and L. Nemuraite [17], M. Mahfoudh and W. Jaziri [18] to cite but a few.

2.2 Designing the DW from a RDB

According to founding father W. H. Inmon, a Data Warehouse is a collection of thematic, integrated, non-volatile, historized and organized data for decision-making [19]. A Data Warehouse (data mart) is a critical component for designing a decision support system. Kimball et al. [23] have introduced the first methodology for designing the logic scheme of a data warehouse. However, the process of this design requires expertise and is too tedious and costly in terms of effort, resources and time. To overcome the complexity and heaviness of this design, several methodologies for automating this design process have been proposed in the literature such as the work of Phipps and Davis [20], Song and al. [21], Mior and al [22] and J. Feki and al. [24].

Approaches to multidimensional modeling, according to Winter et al. [2], can be classified into three classes: demand-driven, supply-driven and hybrid approaches of the two previous classes. The approaches that form part of the 'demand-driven' class begin with the determination of the specific needs of the end-users of the multidimensional model. Contrariwise, the approaches of the supply-driven class begin with the detailed analysis of the data sources in order to determine the concepts of the multidimensional

model. However, hybrid begin with the determination of multidimensional concepts while confronting them with the specific needs of end users. In this work, the hybrid class is adopted.

3 Architectural Model

Our work proposes a multi-approachmodel for semi-automating the design of a Data Warehouse from an ontology. Each integration of a database design method from an ontology to a design approach of a Data Warehouse from a database constitutes a new approach for designing a Data Warehouse from an ontology. Fig.1 presents the overall scheme of the proposed model. Our proposal sets up a model that is an intermediate infrastructure between an ontology (data source) and the design of a corresponding Data Warehouse.

The architectural model of our approach brings together a multitude of phases, each of which is dedicated to a specific task of the design process. Indeed, as illustrated in figure 2, this infrastructure can be divided mainly into the following phases: A) acquisition of the ontology B) design of the corresponding relational databases to the ontology C) implementation of the resulting database within a RDBMS, D) design of the corresponding Data Warehouse and E) implementation of this Data Warehouse within a DBMS. In the multidimensional domain, there are several implementation approaches such as ROLAP, MOLAP, HOLAP, Web-based OLAP, DOLAP, and Real-Time OLAP. For our model, we will use the ROLAP approach by implementing the multidimensional schema within a relational environment. In accordance with this approach, facts and dimensions are materialized by relational tables.

3.1 Ontology Acquisition

It is a preliminary phase whose aim is to make the ontology, which is considered as a source of data and subject of analysis, available to the model. Knowing that an ontology is composed of two parts: Terminological-Box (T-Box) and Assertion-Box (A-Box) [25]. T-Box corresponds to the conceptual model of an ontology representing the concepts linked together by semantic relations. To each concept are attached specific properties, which uniquely identifies it in a domain [5]. A-Box regroups all the instances of the concepts in addition to the instances of the relations as well as the values of the properties [25].

3.2 Relational Database Design

The interest of going through the relational model is to give the designer the hand to closely control the transformation and intervene in the process. Indeed, the model provides the designer with the ability:

- To analyze the data,
- To verify the validity of the data,
- To filter the data,
- To select the useful data,
- To delete the uselessdata,
- To ensure the coherence of the data,
- To ensure the integrity of the data,

In addition, through the relational model, the designer is able to act on the model itself. Consequently, it can add or remove entities, attributes, and relationships.

In fact, the relational model is characterized by its proven capabilities to cope with large amounts of data.

Finally, the passage through relational model allows us to get rid of the inherent complexity of ontologies.

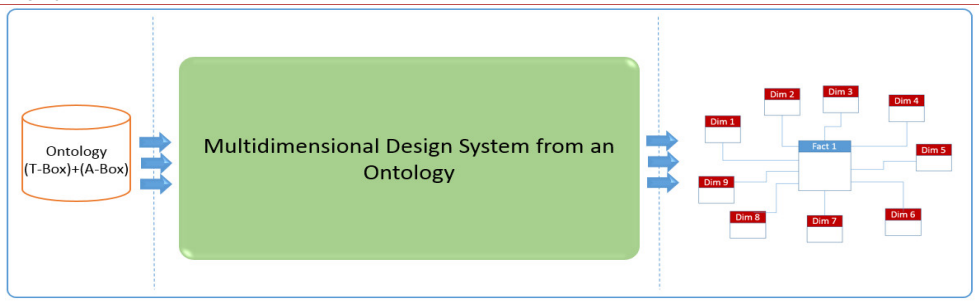


Figure 1. General architecture of a model

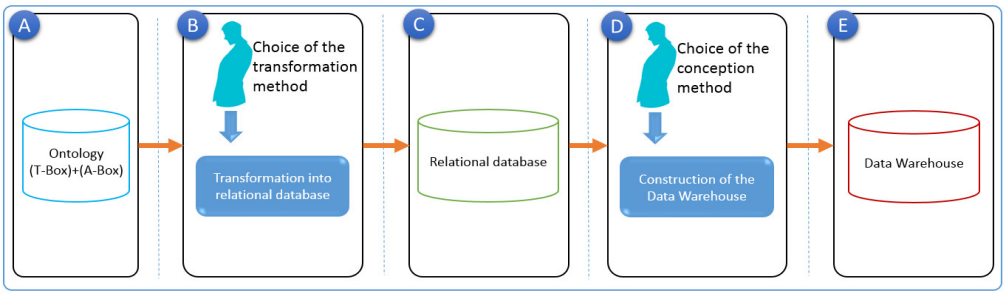


Figure 2. Proposed model architecture

The use of the relational model has substantial advantages, namely maturity, performance, robustness, scalability, reliability and availability [9]. This phase aims at designing a relational database from an ontology taking into account the specific approach chosen by the designer among those to implement in the context of our model.

Several implementations of several approaches to build a relational database from an ontology will be thereforemade available to the designer. The implementation of each approach requires the development of a corresponding algorithm based on the transformation rules dictated by this approach. Among these rules is the transformation of a class into a table (Rule1 in [13]), a 'Data type' property into a column (Rule2 in [13]), and so on.Generally, any class is transformed into an entity at the conceptual level and a table at the logical level. However, the transformation of the attributes of the classes is more complex requiring two types of transformations: the transformation of attributes of type 'Data type' and the transformation of attributes of type 'Object'. Overall, the transformation of a 'Data type' attribute consists of associating it with an attribute of the corresponding entity (conceptual model) or a column of the corresponding table (logical model). On the other hand, the transformation of an attribute of type 'Object' class is much more complex. It consists in matching to this attribute an association at the conceptual level, an intermediate table, or an integrity constraint of foreign key type to the Logical level.

The intervention of the designer allows the parameterization and control of this phase in order to design the relational database corresponding to the basic ontology. This database constitutes the data source (input) for the next phase of the design process.

Each approach for designing a relational database implemented made available to the designer is an alternative for the design process of this database. Therefore, it constitutes an alternative for the design of the Data Warehouse corresponding to the basic ontology. Each choice made corresponds to a particular

result (relational database). The figure 3 shows details regarding the relational database design from an ontology.

3.3 RDB Implementation within a DBMS

The implementation phase of the relational database within a RDBMS is an intermediate phase.

There are two types of ontology-relational coupling methods depending on the type of generated schemas. The first kind of methods aims to design the Conceptual Data Model (CDM) from an ontology as mentioned by Y. Lv and C. Xie in [26], while the second kind of methods aims to design the Logical Data Model (LDM) from an ontology as stated by I. Astrova, N. Korda and A. Kalj in [13]. In the first case, the generated CDM will be transformed into a LDM. The logic model is then translated into a physical model (PDM: Physical Data Model) according to the chosen RDBMS. The physical model is used to generate the SQL DDL (Data Definition Language) script whose execution produces the physical structure of the relational database within the RDBMS.

This phase promotes independence from the underlying RDBMS since it is based on the CDM or LDM of the generated relational database depending on the method kind chosen by the designer.

3.4 Designing of the Data Warehouse

It is at this stage that the design of the Data Warehouse takes place through an approach chosen by the designer from among the approaches to be implemented. The implementation of a specific approach concretizes the vision of automatically designing a Data Warehouse from a relational database. This automation takes place in several steps. Among them are the search for candidate measures as well as candidate facts, the search for candidate dimensions, the search for hierarchies of dimensions, etc. There are methodologies that are based on a data-driven process by focusing on an in-depth analysis of data sources to derive the data warehouse schema in a reengineering process [3]. However, in order to meet users' needs and expectations, the methodologies integrate an analysis phase to avoid users' frustration by allowing them to analyze the data. These approaches are data-driven and requirement-driven as proposed by J. Feki and .al [24].

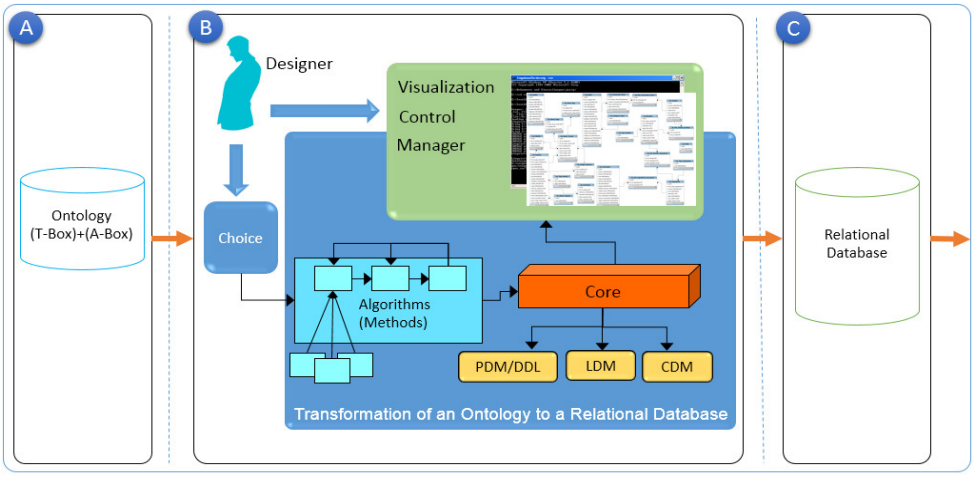


Figure 3. Designing a relational database from an ontolog

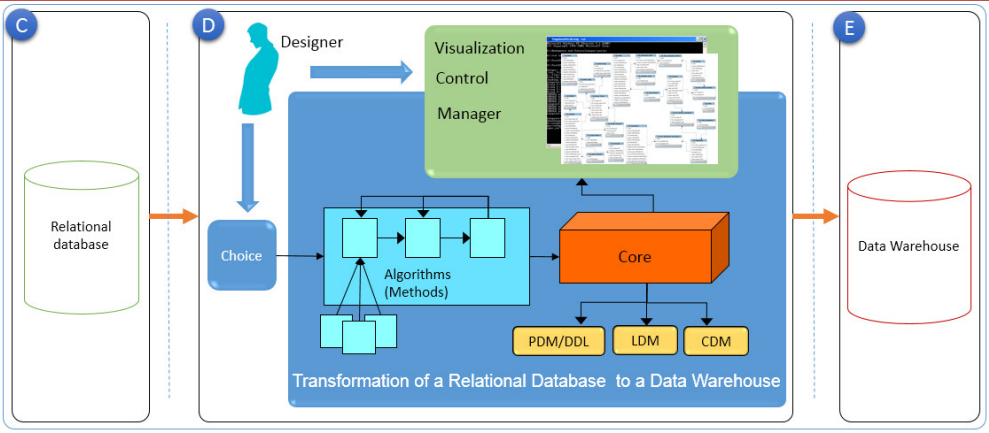


Figure 4. Design of Data Warehouse from a relational database

Figure 4 shows details regarding the Data Warehouse design from a relational database.

By implementing N design approaches of a relational database from an ontology and M design approaches of a DW from a RDB, our model integrates N times M (N * M) Approaches of designing a DW from an ontology thus constituting an extensible multi-approach model (Fig. 8).

4 Example Case Study

As an illustration of our approach in the design of a DW from an ontology, we put forward an example of an application to a domain ontology spelt out by the transitions of the different stages of the design of a corresponding DW. This ontology is relative to the pedagogical domain that we conceived based on a corpus of pedagogical documents, namely the descriptive sheets of the university streams, the results of the students they have obtained in a university stream module, and so on.

Figure. 5 shows a part of the representative graph of this ontology that represents the students and their performances, the university streams and their modular compositions, and so on.

Figure. 6 shows the Entity-Relationship diagram (ERD) of a relational database obtained by applying of the design/transformation rules as dictated by Y. Lv and C. Xie in [26] on the domain ontology. This resulting conceptual schema comprises 7 entities and 7 relationships.

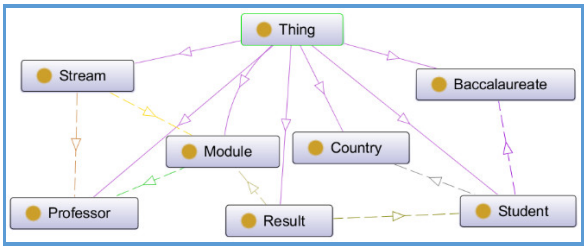


Figure 6. Entity-Relationship Diagram of RDB

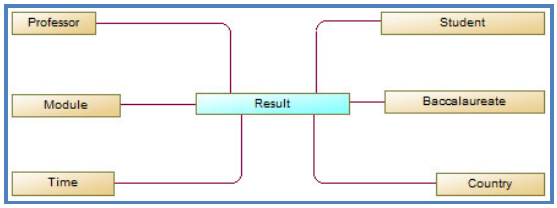


Figure 7. Scheme of Data Warehouse

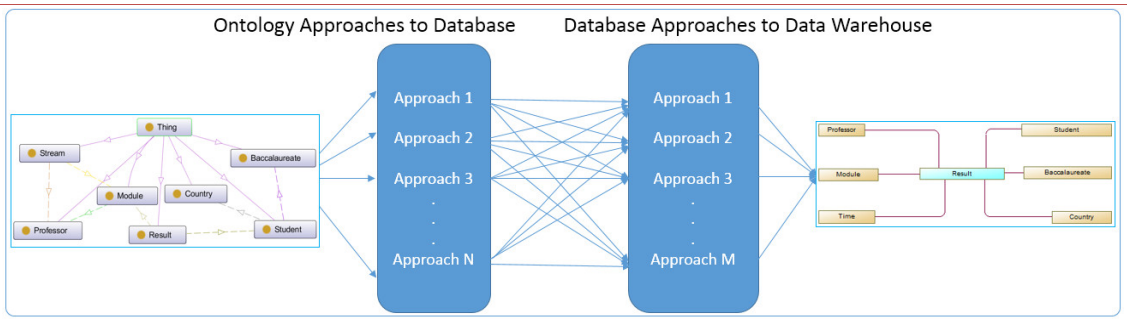


Figure 8. Combination of approaches

By applying the SAMSTAR approach [21] to design a DW from a relational database on the database whose ERD is shown in Figure 7, we identify a candidate scheme with one candidate fact, namely 'Result'. For specific purposes, we keep a scheme centered on the fact 'Result' (because it has highConnection Topology Value) as illustrated in Figure7. This diagram includes in addition several dimensions namely 'Module', 'Student', 'Time', Etc. We add the 'Time' dimensionin accordance with SAMSTAR Algorithm rule 8(iv): Add 'time' dimension to this list.

5 Conclusion

We presented a new multi-approach model to semi-automatize the Data Warehouse design from an ontology. The main objective of the proposed model is to maximize the adequacy of the Data Warehouse designed from an ontology to the end-users business-needs. We have shown that this model provides more ability to adjust the Data Warehouse design from an ontology to the user specific needs. By giving designer the capability to control the process, this model allows to overcome the complexity inherent in ontologies, thus minimizing the gap between data inconsistencies contained in the ontologies and data integrity in the Data Warehouse. In further work, the model should include the expert system in order to further support the designer.

REFERENCES

[1] M. HAJJI, M. QBADOU, K. MANSOURI, "Proposal for a new Systemic Approach of Analitical Processing of Specific Ontology to Documentary resources: Case of Educational Documents", Journal of Theoretical and Applied Information Technology, July 2016, Vol.89, No.2, pp. 481-51.

[2] R. Winter, B. Strauch, "A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects", Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), pp. 231.1, 2003.

[3] O. Romero and A. Abello, "A Survey of Multidimensional Modeling Methodologies," International Journal of Data Warehousing and Mining 5 (2), April 2009, pp. 1-23.

[4] A. Maedche and S. Staab. Kaon: The karlsruhe ontology and semantic web meta project. Künstliche Intelligenz, Special Issue on Semantic Web, Mar 2003.

[5] CHARLET J., BACHIMONT B. & TRONCY R., Ontologies pour le Web Sémantique, in Le Web sémantique, CHARLET J., LAUBLET P. & REYNAUD C. (Ed.), Hors série de la Revue Information - Interaction - Intelligence (I3), 4(1), Cépaduès, Toulouse, 2004, pp. 69-100.

[6] T.R. GRUBER, "A translation approach to portable ontology specifications," in Knowledge Acquisition Journal, 5(2), Academic Press, 1993, pp. 199-220.

[7] R. STUDER, V.R. BENJAMINS and D. FENSEL, "Knowledge engineering: principles and methods," in IEEE Transactions on Data and Knowledge Engineering, 25(1&2), 1998, pp.161-197.

[8] W.E. Grosso, H. Eriksson, R.W. Fergerson, J.H. Gennari, S.W. Tu, and M.A. Musen, "Knowledge Modeling at the Millennium (the Design and Evolution of Protege-2000)," Proc. 12th Workshop Knowledge Acquisition, Modeling and Management (KAW '99), 1999.

[9] E. F. Codd. 1990. The Relational Model for Database Management: Version 2. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[10] Dan Brickley and R.V. Guha, "Resource Description Framework (RDF) Schema Specification 1.0. W3C Candidate Recommendation", 2000.

[11] A. Gali, C. Chen, K. Claypool and R. Uceda-Sosa, "From ontology to relational databases," in Conceptual Modeling for Advanced Application Domains, LNCS, vol. 3289, 2005, pp. 278–289

[12] E. Vysniauskas, and L. Nemuraite, "Transforming ontology representation from OWL to relational database," Information Technology and Control, vol. 35A, no. 3, 2006, pp. 333-343.

[13] I. Astrova, N. Korda and A. Kalja, Storing OWL Ontologies in SQL Relational Databases, International Journal of Electrical, Computer and System Engineering, 2007.

[14] C. Fankam, L. Bellatreche, H. Dehainsala, Y. Ait-Ameur and G. Pierra, "SISRO : Conception de bases de données à partir d'ontologies de domaine", TSI Volume 28, pages 1-29, 2009.

[15] J. Trinkunas and O. Vasilecas, "A Graph Oriented Model For Ontology Transformation Into Conceptual Data Model", Information Technology and Control, 2007, Vol. 36(1A), pp. 126–132.

[16] I. Astrova, N. Korda and A. Kalja, "Storing OWL ontologies in SQL Relational Databases," Engineering and Technology, 2007, Vol. 23, pp. 167–172.

[17] E. Vysniauskas and L. Nemuraite, "Mapping of OWL ontology concepts to RDB schemas," in Information Technologies' 2009, Proceedings of the 15th International Conference on In-formation and Software Technologies; 2009. pp. 317-327.

[18] M. Mahfoudh and W. Jaziri, "Approche de couplage de BD et d'ontologie pour l'aide à la décision sémantique : contribution pour la satisfaction des requêtes SQL et SPARQL," Techniques et Sciences Informatiques, volume 32 of Hermes, pages 863-889, 2013.

[19] W.H. Inmon, "Building the Data Warehouse", Second Edition, New York: John Wiley & Sons, 1996.

[20] C. Phipps, K.C. Davis, "Automating data warehouse conceptual schema design and evaluation", Proc. of the International Workshop on Design and Management of Data Warehouses (DMDW'2002), vol. 58, pp. 23-32.

[21] Y. Song, R. Khare, B. Dai, "Samstar: a semi-automated lexical method for generating star schemas from an entity-relationship diagram", in Proceedings of the ACM tenth international workshop on Data warehousing and OLAP, pp. 9-16, 2007.

[22] M. N. M. Nazri, S. A. M. Noah and Z. Hamid, "Automatic data warehouse conceptual design," 2008 International Symposium on Information Technology, Kuala Lumpur, Malaysia, 2008, pp. 1-7.

[23] R. Kimball, L. Reeves, W. Thornthwaite, M. Ross, "The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing" in , 1998, John Wiley & Sons, Inc.

[24] J. Feki, A. Nabli, , H. Ben-Abdallah and F. Gargouri, "An Automatic Data Warehouse Conceptual Design Approach". Encyclopedia of Data Warehousing and Mining, John Wang Edition, 2008.

[25] I. Horrocks, "Ontologies and the semantic web", Commun. ACM, vol. 51, no. 12, pp. 58-67, 2008.

[26] Y. Lv and C. Xie, "An ontology-based approach to build conceptual data model," 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, Sichuan, 2012, pp. 807-810.

[27] T. Podsiadły-Marczykowska, T. Gambin, and R. Zawiślak, “Rule-Based Algorithm Transforming OWL Ontology Into Relational Database,” in Beyond Databases, Architectures, and Structures: 10th International Conference, BDAS 2014, Ustron, Poland, May 27-30, 2014. Proceedings, S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, and D. Kostrzewa, Eds. Cham: Springer International Publishing, 2014, pp. 148–159.

[28] V. Nebot and R. Berlanga, “Building data warehouses with semantic web data,” Decis. Support Syst., vol. 52, no. 4, pp. 853–868, 2012.

[29] D. Colazzo, I. Manolescu, A. Roatis, and A. Roati, “Warehousing RDF Graphs To cite this version : Warehousing RDF Graphs ∗,” Hal, 2013.

[30] D. Colazzo, I. Manolescu, A. Roatis, and A. Roati, “Warehousing RDF Graphs To cite this version : Warehousing RDF Graphs ∗,” Hal, 2013.A.

[31] B. B and A. Bagnall, “A Value-Added Approach to Design BI Applications,” vol. 9263, pp. 257–269, 2015.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Sentiment Analysis tool for Pharmaceutical Industry & Healthcare

[1]Hanane Grissette, [1]EL Habib Nfaoui, [2]Adil Bahir
[1]LIIAN Laboratory, Sidi Mohamed ben Abdellah University
Fez Morocco.
[2]IT Head/Business Information Manager,
Novartis pharma, Casablanca Morocco.
Hanane.grissette@usmba.ac.ma; Elhabib.nfaoui@usmba.ac.ma; adil.bahir@novartis.com

ABSTRACT

Sentiment analysis (SA) is broadly used to analyze people's opinions about a product or an event to identify breakpoints in public opinion. Particularly, pharmaceutical companies use SA to ensure they gain a competitive edge through better understanding of patients' experiences allowing for more personalization and high responsiveness to consumers on social media. Patients self-reports on social media, frequently capture varied elements ranging from medical issues, product accessibility issues to potential side effects. The exploitation of such overwhelming unstructured data on social media through SA is of critical importance, however, the general-purpose sentiment analysis tools are not adapted and do not incorporate the specific lexicon used in the life sciences/pharma context which reduces the ability of such tools to accurately detect the meanings of the sentiments expressed towards treatments/scientific studies and pharma companies at large. Indeed, those tools involve generalized dictionaries, techniques and methods that face several challenges to detecting the suitable sentiment polarity. In this paper, we develop a dedicated research tool that extract in real time the sentiments and emotions conveyed by users on social media regarding pharmaceutical industries. Our focus is to improve the analyze of opinion about a target in pharmaceutical industries based on hybrid approaches used in single unified system throughout sentiment analysis process to seek the relevant sentiment polarity, which is tailored specifically to detect positive, negative or neutral opinions of patients and consumer health. This paper proposes a first architecture validated by a real and industry case.

Keywords: Pharmaceutical industry, Sentiment Analysis, real-time analytics, healthcare, public mood,Social media.

1 Introduction

Sentiment analysis (SA) has received great attention recently, due to the huge amount of user-generated information onsocial Media, which deals with the computational treatment of opinion, sentiment and other components. Mining social media presents a direct response to the surge of interest in understanding the conversations, identifying the relevant content and acting appropriately on it to identify new opportunities, deal with new business solutions and manage reputations. The Pharmaceutical Industry is no exception and SA is changing the game for Pharma companies in terms of

DOI: 10.14738/tmlai.54.3339
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3339

understanding the sentiments and the conversation around their products and is becoming a key pillar in their digital strategy. TABLE I illustrates key highlights and issues faced by patients or HCPs that emerged through posts on Facebook & Twitter.

Several recent publications assume to the richness of information to be found in patient self-reports of their problems in social media, and the useful reports are enhanced, thus absolutely aiding pharmaceutical industries to get patient voice and capture unpreventable issues.

Although there are several studies and tools of SA, few tools in the field of pharmaceutical industries use SA to analyze opinion considering the person himself or other entities as target of analysis[1]. Thus, the performance of SA tools in this field is mainly related with polarity classification task, when bad scoring of terms can effectively affect the global sentiment polarity, that is occurs when a piece of text stating an opinion on a single issue that can be classified as one of two opposing sentiments regarding to a specific context. However, according to specific industry, the target data to be analyzed can have a different distribution than the data-driven dictionaries offered. Indeed, polarity classification must identify professional and colloquial expressions and terms based on both dictionaries and machine learning techniques to improve the quality of detecting the suitable sentiment polarity, making the product evaluations more credible. Even though, the accurate meaning of expressions based on context which it appears that has great importance in determining the relevant polarity of sentiment.

Recognizing the importance of mining social media using SA techniques for pharmaceutical industries, we are motivated to develop a dedicated research tool that improve the detection of positive or negative people's opinions regarding pharmaceutical industriesin real time, by inspecting posts of authors, medical entities and posts about events. This tool involves several APIs, that's based on statistical learning techniques and lexicon based approaches.

Table 1. Examples of Patients Posts On Social Media

PostID	Post content	Type of post	Report source
844476 028055 703552	Woke up too late to take my **Concerta [Drug]** yesterday so I'm much just slept through the entire day and only woke up an ho	Patient Experience	Cancer patient
844247 270870 716418	Where do people get **Concerta [drug]?** hey i have **inattentive adhd** and im on meds (concerta) but im having a problem about...	Drug/Product Accessibility issue	Patient
843824 219754 381312	evr to be clear...I'm on Concerta [drug] and in the official name it has '**Methylphenidate**'. It's a legal form of it....	Patient Experience	Hearsay
844536 299726 032896	**Novartis [Pharmaceutical company]** heart-failure med #Entresto cuts **Hb A1c,**insulin starts in diabetics | **FiercePharma**....	HCP feedback	Diabetes
844525 854600 290304	**Novartis [Pharmaceutical company]** heart drug portfolio hit by failed **serelaxin**[**investigational drug]** study – Reuters.	Event	Fitness news magazine

The remainder of this paper is organized as follows, Section 2 presents andsummarizes related works. Section 3 describes SA process. Section 4 describes our proposed tool. Section 5 presents results, discussions, limits and perspective as research questions in SA field. Section 6 concludes this paper.

2 Related Work & Applications

SA becomeinterested part of real-life applications in various areas like business intelligence, market prediction, Recommender System, Marketing intelligence, attrition prediction.

Some works were done about SA, especially comparing the tools proposed. The work presented in [2] makes the comparison of nine SA tools: AchemyAPI, Lymbix, MLAnalyzer, Repustate, Semantria, Sentigem, Skytle, Textalytics and Textprocessing. To calculate the accuracy of each tool, texts of different sources were collected (news, comments and tweets). The tools with the greatest accuracy were Textalytics (75%), Skytle (73%) and Semantria (68%).

In other work, [3], SA tools were also compared. Twenty tools

were chosen: fifteen stand-alone SA tools (SentiStrength, Chatterbox, Sentiment140, Textalytics, Intridea, AiApplied, ViralHeat, Lymbix, SentimentAnalyzer, TextProcessing, Semantria, uClassify, MLAnalyzer, Repustate and a last one referred to as Anonymous by the authors) and five workbench tools (BPEF, Lightside, FRN, EWGA, RapidMiner). The texts used to analyze the tools were tweets related to the themes: telecommunications, pharmaceutical, security, technology and consumer products at retail. Among the stand-alone tools, the one with the greatest average accuracy (67%) was the SentiStrength. Among the workbench tools, BPEF presented best average accuray (71%).Another study is Ref. [4]. This tested AlchemyAPI, OpenAmplify and Texterra on public text. The last one was a new tool proposed by the article and presented accuracy of 79%, higher than the OpenAmplify (57%) and AlchemyAPI (42%).

Another Study is Ref. [5]. Were conducted on existing SA methods, they measure the percentage of agreement for each method from the messages classified for each pair of methods, for what fraction of these messages they agree. They find that some methods have a high degree of overlap asin the case of LIWC and PANAS-t (80%), while others have very low overlap such as SenticNet and PANAS-t (30.77%). PANAS-t and Happiness Index had no intersection. In summary, the above results indicate that existing tools vary widely in terms of agreement about the predicted polarity, with scores ranging from 33% to 80%. This implies that the same social media text, when analyzed with different sentiment tools, could be interpreted very differently. In particular, for those tools that have lower than 50% agreement, the polarity will even change accordingly to pharmaceutical industries (e.g., from positive to negative, or vice versa). Nevertheless, severalstartupsthatmeasureopinionsonrealdatahaveemergedanda numberofinnovativeproductsrelatedtothisthemehavebeendeveloped.Thereare multiplemethodsformeasuringsentiments,including lexical-basedandsupervised machinelearningmethods[6]. Despitethevastinterestonthethemeandwide popularityofsomemethods,itisunclearwhichoneisbetterforidentifyingthe polarity for posts. Indeed, pharmaceutical, medical professionals and health vocabularies are often observed that are different from those of public[7]. To interpret medical terms in online patient discussions, Consumer Health Vocabulary (CHV), a lexicon linking UMLS standard medical terms to patients' colloquial language [8], is adopted in few recent studies.We summarize inTABLE II the recent work of SA in health and pharmaceutical field. We have selected three criterions: methods, approaches and data source.

3 Sentiment Analysis Process

In this section, we describe the broad dimensions and the process of SA. As defined in [19], sentiment analysis is a computational study of opinions, sentiments, Subjectivity, emotions, and attitude expressed in texts towards an entity. Thus, SA is a multi-task of detecting, extracting and classifying opinions, sentiments and attitudes concerning different topics, as expressed in textual input.In Fig 1.presents seven broad dimensions of SA.

Table 2: Works Summary of Related Work In Pharmaceutical & Healthcare

Prior Study	Description	Data Source/Methods/approaches
Benton et al. [9] (published in 2011)	They proposed advanced techniques for analyzing health social media content.	They adopted three breast cancer forums as test bed. Using Lexicons: CHV, FAERS for medical entity Extraction approaches.
Mao et al. [10] (published in 2013)	In this work, the proposed architecture Help to understand drug switching and discontinuation behaviors to understand patient reported adverse drug events.	They developed their test bed on 12 breast cancer forums. Using Lexicons: CHV, FAERS for medical entity Extraction approaches.
Sarker et al. [11] (published in 2015)	The proposed method improves detection of sentences with ADE mentions, they developed a text classification method for adverse drug reaction detection on clinical reports.	Data source is obtained from clinical reports, Twitter, and daily strength. Using Lexicons: UMLS, WordNet, MedEffect, SIDER, and COSTART.
Ramon et al. [1] (published in 2016)	This study present a Sentiment Analysis tool, named SentiHealth-Cancer (SHC-pt) that improves the detection of emotional state of patients in Brazilian online cancer communities, by inspecting their posts written in Portuguese language.	The SHC-pt tool is based on the method called SentiHealth proposed in ther paper [19]. Moreover, as a second alternative is to analyze the texts in Portugues.
Segura-Bedmar et al. [12] (published in 2015)	They developed a distant supervision approach to extract adverse drug events in Spanish	Data is collected from Forums and blogs, they used GATE pipeline for medical entity recognition and distant supervision with shallow linguistic kernel for adverse drug event extraction.
Chee et al. [13] (published in 2011)	They developed ensemble classifiers with SVM and nave Bayes to classify risky drugs and safe drugs based upon online discussions	Data is collected from Health forums in Yahoo! Groups; SVM and Naive Bayes are used for classification tasks; Lexicons: UMLS, MedEffect, SIDER for medical entity recognition approaches.
Wu et al. [14] (published in 2016)	They developed a discriminant classifier with Ricchio method and a generative model to determine whether the side effect is relevant to the drug	Data Source is collected from online discussion;Rocchio method is used for classification task.
Nikfarjam et Gonzalez [15] (published in 2011)	They proposed a method to capture the underlying syntactic and semantic patterns from the Daily Strength reviews.	Reviews are collected from social media e.g. “Daily strength”; association rule mining is used for entity recognition task & approaches to generate patterns for adverse event recognition. This study benefited from a large, manually annotated corpus (1,200 records) and an algorithm that can detect expressions not included in a lexicon.

K. Portier et all [16] (published in 2013)	The objective of this study is understanding topics and sentiment, detect pessimistic emotional state. And also detect changes in a person mood as consequence of his interactions with other patients in a community. That is used to give emotional support to patients.	SA techniques applied on posts in cancer online communities.
Liu et all .[17] (published in 2011)	They are classifying Yahoo Answers posts into patients' posts or medical professionals' posts 82.0% in f-measure for identifying drug user from tweets.	They used test beds such as Twitter, hashtags and URLs are adopted in the feature set. They used sentiment related words from LIWC as features in classification. In terms of learning methods, the most commonly employed learning algorithms are Support Vector Machines (SVM) and naive Bayes.
Abdul-Mageed et al. [18](published in 2014)	This study Performed subjectivity and sentiment analysis (SSA) of social media for a morphologically-rich language.	They Worked on 2798 chat turns, 3015 Arabic tweets, 3008 sentences from 30 modern standard Arabic Wikipedia Talk pages, and 3097 web forum sentences.

3.1 Subjectivity Classification

Subjectivity classification is the property associated with words and word sense that deals with the detection of ''private states''– a term that encloses sentiment, opinions, emotions, evaluations, beliefs and speculations. To deal with opinion word (features) and product aspect extraction considered named entity recognition (NER) system for streaming micro-text. It was based on different contextual associations like semantic association, temporal association, and social association between micro-text clusters.

3.2 Sentiment Classification:

Sentiment classification is the determination of orientation of sentiment of given text in two or more classes.

3.3 Polarity determination

Sentiment classification is concerned with determining polarity of a sentence, whether a sentence is expressing positive, negative or neutral sentiment towards the subject. Hence, Sentiment classification is also termed as polarity determination.

3.4 Opinion feature extraction and product aspects extraction

On social networking, if someone is linked to/followed/liked a social networking group or community or public figure, it implies that the person has either positive or negative feelings

Towards that entity. To perform finer grained SA regarding to aspect of product and opinion word as features, we need to identify feature's opinion and its Sentiment score of different aspect that will affect global sentiment strength.

Otherwise, Sentiment analysis is not a single problem, various steps are needed to perform Sentiment Analysis from given texts, are presented inFig 2.

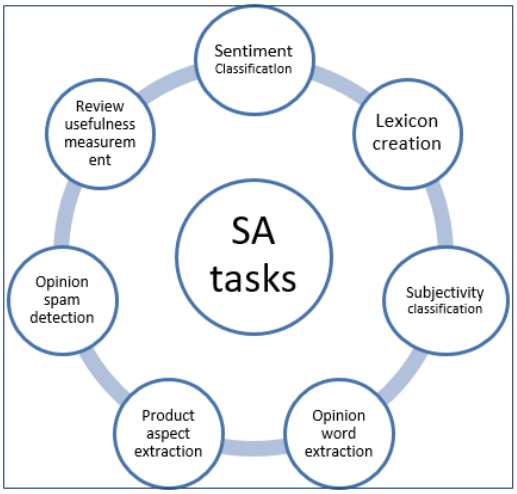


Figure 1: the seven broad dimensions of SA

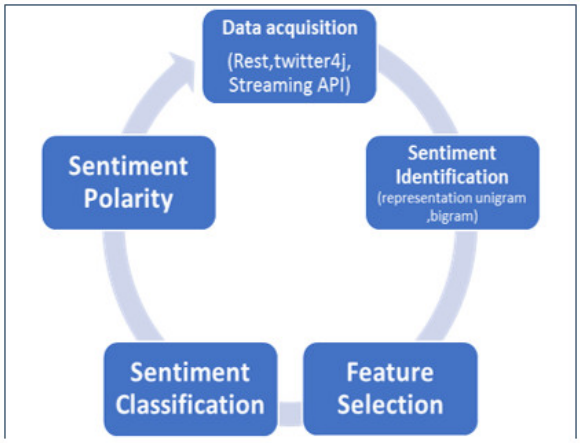


Figure 2: Sentiment analysis steps

SA techniques can be roughly divided into machine learning approach, lexicon based approach and hybrid approach. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a sentiment lexicon (a collection of known and precompiled sentiment terms), that's divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in most methods.

4 Proposed Architecture &Process

The objective of this work is to present a SA tool, that improves the detection of opinions regarding pharmaceutical industry in real time andanalyse of emotional state of peoples talking about, to get up to the minute information on social media.That isexplained in detail belowFig 3.

4.1 Pharmaceutical industry data collection:

There are several social networks that people can use to interact with each other. The most widely used of this are Facebook and twitter. In January 2017, the network contained 1.86 billion monthly active Facebook users, where an active user is one who has realized log-in in the last 28 days and also has at least one fried added to his network. From a recent study, 54% of patients are very comfortable with their

providers seeking advice from social networks to better treat their conditions. In other study, more than 40% of consumers say that information found via social media affects the way they deal with their health.

Another advantage of using Facebook and twitter are these they have a lot of patients and they are very active considering the content disseminated in them. For that, we developed an automated crawler program to collect data from Facebook & Twitter. By inspecting opinions of pharmaceutical industries, our collector program is based-keyword to seek opinions about specific products, consumer health and events in the pharma world. Collected data includes post ID, URL, post content, post author's ID and post date.

As our study focuses on sentence level information extraction and processing, we handle our source data obtained from crawler program in table and parse post content to be an input of next step of analysis.

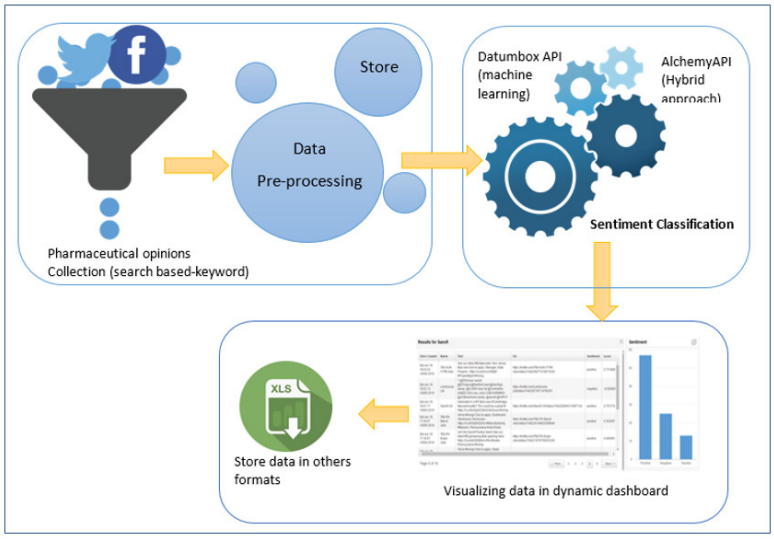


Figure 3: Major Components of proposed architecture

4.2 Sentiment Classification:

It is a challenging task to classify posts from noisy patient-generated content and professional, according to medical and pharmaceutical expressions. Lexicon based approach was the best performing medical entity recognition system in prior studies [20]. Our tool incorporates the statistical learning methods and machine learning approaches also known as hybrid approaches to identify the suitable sentiment polarity. Indeed, proposed tool involves methods from Alchemy API & Datumbox. When AlchemyAPIincorporatesstatistical analysis, techniques based on machine learning algorithms, and linguistic techniques from natural language processing into a single unified system. It support to English and other languages and can be used by applications programmed with java, C/C++, C#, Perl, Java, PHP, Python, JavaScript, Ruby and can also be used by android applications [51]. A free version is available that allows to perform up to a thousand daily transactions and five concurrent request. In addition, Datumbox API is a machine learning Framework presents a several Text Classifiers such as the Naive Bayes, the SoftMax Regression and the Max Entropy used to classify Sentiment based on statistics methods.

Through parameters, to perform SA on document level (reviews and posts are considered as document). We present the suitable sentiment polarity of each document on dynamic dashboard. The results of the analysis may be output in excel formats, to deal better with other purpose.

5 Case Study

5.1 Novartis international company in Maghreb

This work is part of a project evaluated and approved by industry case, one of the global leader in the pharmaceutical industry, Novartis International AG in Morocco country that is a Swiss multinational pharmaceutical company.With the growing availability and popularity of opinion-rich resources such as social media. Table III presents a summary of search volumes for given keyword “Novartis” to assess the potential search engine traffic for Novartis brand on the web.

Table 3: Daily Volume of Search Keyword Novartis from Wordtracker

keyword	Volume
novartis	**1,428**
Novartis Emcurevysov	**108**
Novartis Emcure partnership	**106**
Novartis and vysov	**104**
Novartis usv partnership	**86**
Novartis ophthalmology	**65**
Novartis rebate	**56**
Novartis osteoporosis	**41**

Otherwise, our statistics performed on Facebook and twitter presented inFig4 that both show the strong presence represented by sum of Facebook fans and Twitter followers in different pages.Other statistics track user interactions on dedicated pages of pharmaceutical industry on Facebook and Twitter respectively that are presented in Fig 5. And Fig 6.Nevertheless, Novartis pharma industry have very motivated to poll these microblogs to get insight from people’s experiences and related-drug problems of their product and performed events… etc.

There are few works proposing context-domain SA solutions aiming at improving accuracy of the classification of opinions in the pharmaceutical context that involves driven-domain methods and approaches to get right sentiment context. Based on the research gaps, to adapt generalized-purpose SA tools at local context, that is identified by many industries cases. For that, we are motivated to adapt our tool at Novartis pharma context, the final objective is to get insights of minute information about pharmaceutical industries on social media, whether its Novartis pharma, its products, its event or its competitors in the world.

5.2 Validation and Experimental Result

In this section, we present the results obtained by applying tool on the different collections of data collected from Facebook & twitter. The conducted experiments consider the language of posts are written, the most of analyzed texts are in English language.(For exampleFig 7).

Figure 7: tweet of Ritaline [Drug] detected in English language.

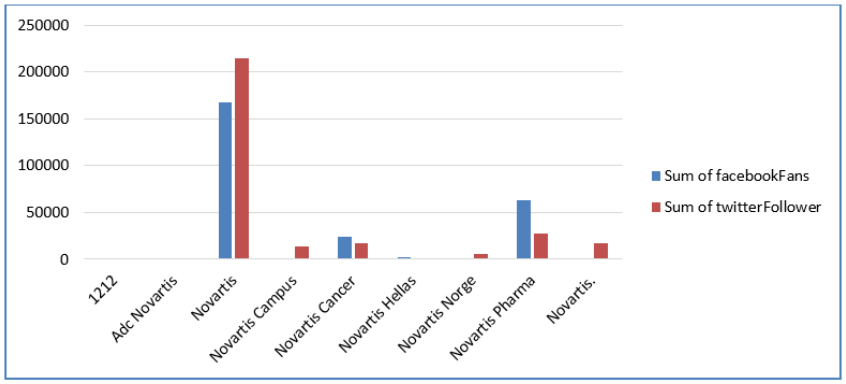


Figure 4: sum of Facebook fans and Twitter followers in different pages

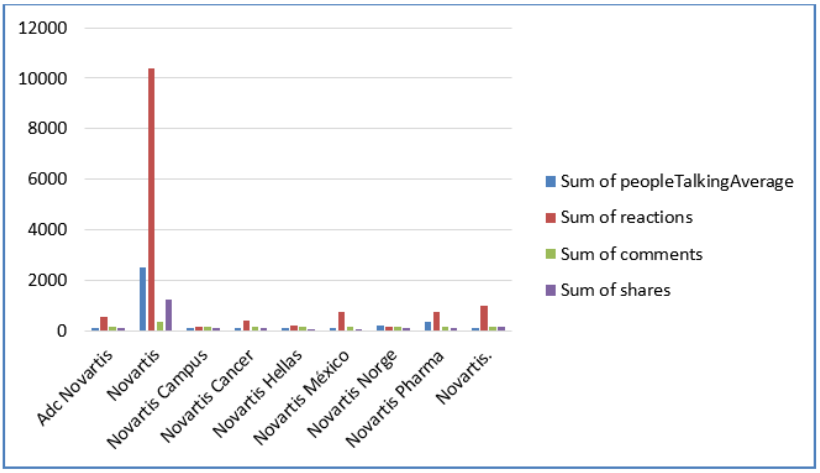


Figure 5: interaction rate by weekday on all dedicated Novartis pages

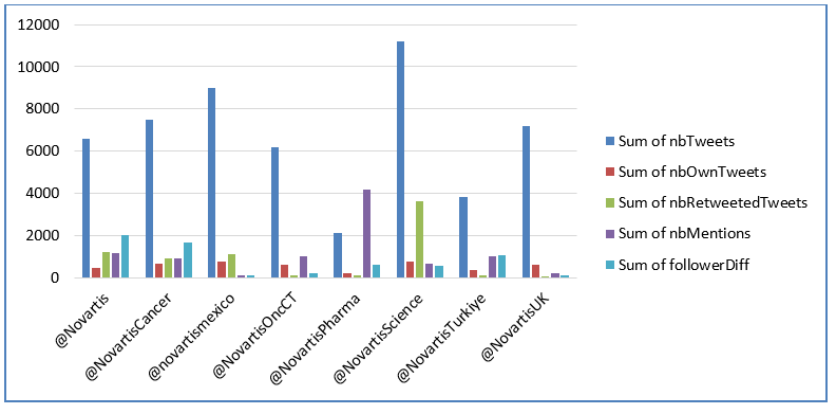


Figure 6: average Followers on dedicated Novartis pages in twitter and its interactions

Facebook and twitter are the most popular platform used by patients and professionals to share their experiences. For that, we conduct our experiments to extracting people-report on Novartis products & events, our crawler program randomly selected 100 posts in real time based on keyword. The result presented in TABLE Vsummarizes the results of some experiments performed on Facebook pages in real time. In addition. As shown in Fig 8. The 100 post collected considering "Novartis" opinions (the crawler program must collected all tweet contains it), 30 post are positive, 55 are neutral and 15 are negative. Others results presented bellow shows there is a mix of sentiment in the collected post.

Table V: Results Obtained of the Experiments Facebook

POSTS DATE	KEYWORD USED	POSITIVE	NEGATIVE	NEUTRAL
FRI ABRIL 18 13:42:01	Novartis	30	15	55
FRI ABRIL 18 13:48:01	Novartis pharma	32	12	65
FRI ABRIL 18 13:50:01	Concerta [dug]	7	5	88
FRI ABRIL 18 14:42:01	Novartis+ Sanofi [pharmaceuti cal company]	38	22	40

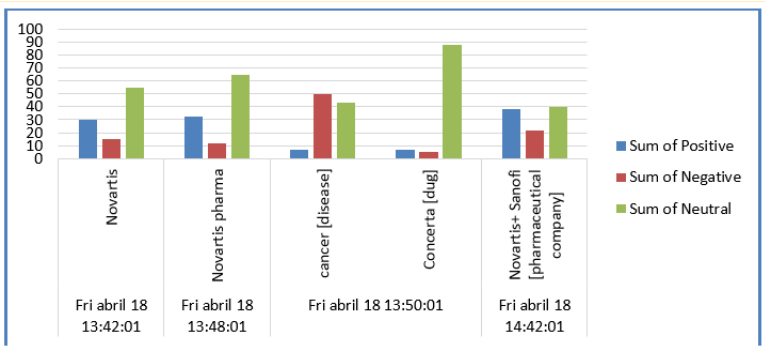


Figure 8: average of sentiment polarity of opinions in Facebook pages

Table 4: Results Obtained of the Experiments on Twitter

POSTS DATE	KEYWORD USED	POSITIVE	NEGATIVE	NEUTRAL
SAT JUN 18 13:38:01	Novartis	63	17	20
FRI ABRIL 18 13:48:01	Novartis pharma	35	50	15
SAT JUN 18 13:50:01	Alcon[Novartis Filial]	42	28	30
SAT JUN 18 15:42:01	Sanofi[American pharmaceutical company]	57	25	18
SAT JUN 18 15:43:01	Pfizer[American pharmaceutical company]	58	25	17
SAT JUN 18 14:45:01	GSK[British pharmaceutical company]	50	23	27
SAT JUN 18 10:24:01	Concerta [dug]	19	60	21
SAT JUN 18 10:22:01	Exelon[dug]	42	17	41
SAT JUN 19 10:32:01	Jakavi[dug]	5	1	1
SAT JUN 19 10:42:01	Ritalin[dug]	40	45	15

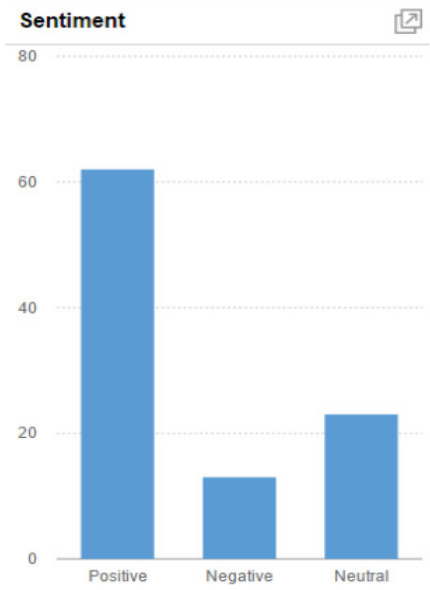


Figure 9: Screenshot of dynamic dashboard global polarity sentiment obtainedby tool

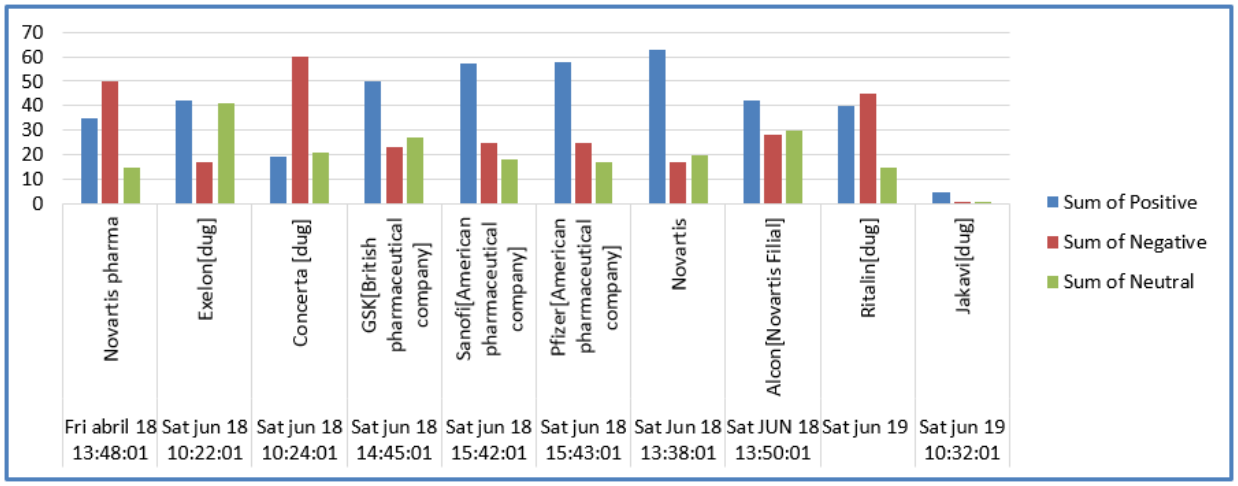
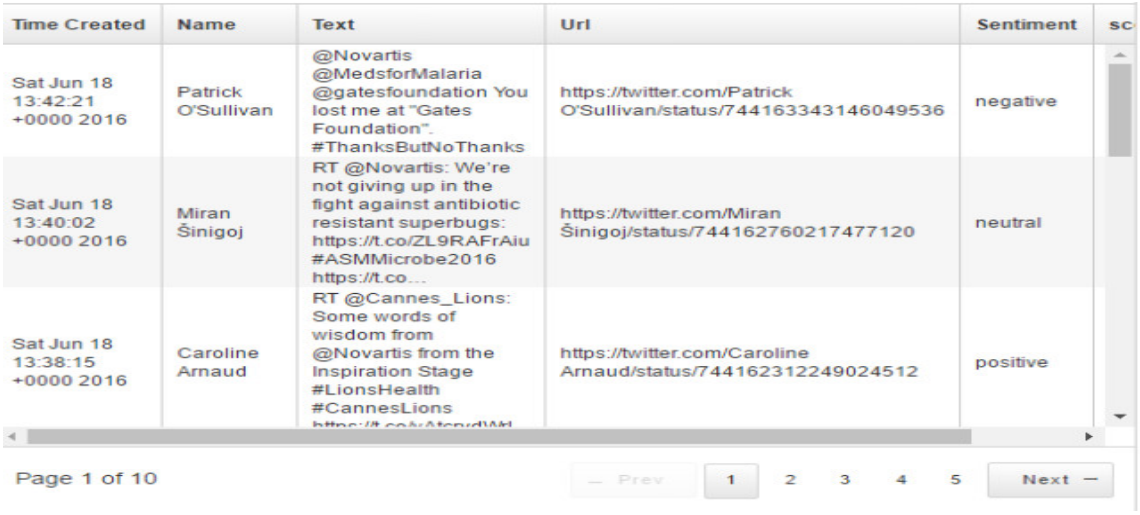


Figure 10: summarize some people's opinions on twitter at 18/19 June in 2016

Regarding the context of pharmaceutical & health, many people usually choose twitter to share their problem or to search information about drug-treatment, for that we conduct another experiments to hear people's voice on twitter,TABLE IVpresent some results obtained at 18 in June 2016. From Fig 10. It's clear that on twitter people are more interactive considering twitter as credible microblogs to share and change health and treatment information.

Fig 11. Showeda screenshot of data collected by our crawled program and their classification obtained in our prototype, which is based on different keyword to seek opinions about them. Accompanied with a dynamic dashboard to visualize results that show the global sentiment polarity of opinion Fig 9.



Time Created	Name	Text	Url	Sentiment	sc
Sat Jun 18 13:42:21 +0000 2016	Patrick O'Sullivan	@Novartis @MedsforMalaria @gatesfoundation You lost me at "Gates Foundation". #ThanksButNoThanks	https://twitter.com/Patrick O'Sullivan/status/744163343146049536	negative	
Sat Jun 18 13:40:02 +0000 2016	Miran Šinigoj	RT @Novartis: We're not giving up in the fight against antibiotic resistant superbugs: https://t.co/ZL9RAFrAiu #ASMMicrobe2016 https://t.co...	https://twitter.com/Miran Šinigoj/status/744162760217477120	neutral	
Sat Jun 18 13:38:15 +0000 2016	Caroline Arnaud	RT @Cannes_Lions: Some words of wisdom from @Novartis from the Inspiration Stage #LionsHealth #CannesLions	https://twitter.com/Caroline Arnaud/status/744162312249024512	positive	

Page 1 of 10 — Prev 1 2 3 4 5 Next —

Figure 11: Screenshot of tweets dataobtained by tool and their classification obtained.

The accuracy is mainly based on the quality of the annotation, word's score, and recognition entity. In order to better adopt this tools in a specific domain, usually the training process will take a long time to summarize their customized entity of each domain and term at all.

Besides, the performance of algorithms and techniques of sentiment analysis is usually not good when oriented to a local context or domain such as score of entities changed in some context.Also, Such Brand

or Industry has its knowledge base and its concepts, especially when we spoke about Pharmaceutical industry. There is word's polarity change from context which it appears to another.Therefore, the meaning and context's terms used in posts about product's Novartis can affect global polarity. For example, the tweet at Fig12 below has been classified as negative tweet by AlchemyAPI, but tweet present a very important result offered by Novartis study, so it's a positive tweet for Novartis. Even if it talks about patients with the most serious form of childhood arthritis.

We notice that the same gaps exist in another domain. For example, in [20] an automobile organizationhave a same gap of analysis, that is concerned only aboutrecognizing the sentiment information received forautomobiles only. Therefore, a list of Domain Dedicated PolarWords (DDPW), proved as the best lexical resource fordomain oriented sentiment analysis.

The research work presented in this paper allows to analyze the sentiment expressed towards pharmaceutical products according to local context. Therefore, as a future extension, we will propose a dynamic method to recognize semantic meaning according to local context, regarding to medical entity, persons, and drug's name, adverse drug event (ADE) and others features. In this context, some research question is remaining to resolve:

- What are the requirements and methodologies for developing extensible and scalable SA Model in real time?
- How statistical learning techniques can allows to create dynamic dictionaries regarding pharmaceutical and medical entities?
- What are considerations should be taken into account for improve the quality of SA according to local context?

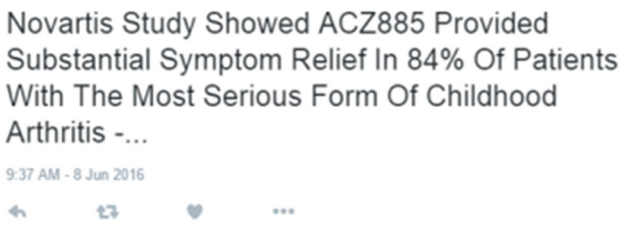


Figure 12. Tweet about patients with the most serious form of childhood arthritis.

AKNOWLEDGMENT

Thanks to Novartis Pharma Company in Morocco for partnering and supporting the present research.

6 Conclusion

The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world.

In the proposed tool, we classified sentiments and emotions conveyed by users in real time as positive, negative and neutral, regarding the pharmaceutical industry. Many experiments were done towards different Novartis's product and its competitors. It would be interesting to extend this work to consider more features that improve the quality of classification of sentiments and recognition of entities. This makes possible the SA establish other polarities that show more clearly personal emotions.

REFERENCES

[1] Ramon et al. "SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks", Published by Elsevier Ireland, http://dx.doi.org/10.1016/j.ijmedinf.2015.09.007.

[2] Cieliebak et al. , "Potential and limitations of commercial sentiment detection tools", Proceedings of the First International Workshop on Emotion and Sentiment in Social and Expressive Media: approaches and perspectives from AI (ESSEM 2013), A Workshop of the XIII International Conference of the Italian Association for Artificial Intelligence (AI*IA 2013), Turin, Italy, December 3, 2013, 2013, pp. 47–58 http://ceur-ws.org/Vol-1096/paper4.pdf.

[3] Abbasi et al. "Benchmarking twitter sentiment analysis tools",Proceedings of the Ninth International Conference on Language Resourcesand Evaluation (LREC-2014), Reykjavik, Iceland, May 26–31,2014,2014,pp.823–829http://www.lrec-.org/proceedings/lrec2014/ summaries/483.html.

[4] Turdakov et al." aframework for text analysis", Program. Comput. Software 40 (5) (2014)288–295,http://dx.doi.org/10.1134/s0361768814 050090.

[5] Pollyanna et al. "Comparing and Combining Sentiment Analysis Methods", not published.

[6] Montoyo et al., " subjectifivty and sentiment analysis:: an overview of the current state of the area and envisaged developments", decis support syst.53 (2012) 675–679.

[7] KANG et al. " review-based measurement of customer satifaction in mobile service: Sentiement analysis and vikor approach, expert syst. APPL.(2013), HTTP://DX.DOI.ORG/10.1016/J.ESWA.2013.07.101

[8] KUHN et al., "A side effect resource to capture phenotypic effects of drugs", MOL. SYST. BIOL. 6 (1) (2010) 343.

[9] BENTON et al., "Identifying potential adverse effects unsing the web: a new approach to medical hypothesis generation", J. BIOMED. INFORM. 44(6) (2011) 989–996

[10] MAO et al. " Online discussion of drug side effects and discontinuation among breast cancer survivors, Pharmaceopidemiol. DRUG SAF".22 (3) (2013) 256–262.

[11] SARKER et al." portable automatic text classification for adverse drug reaction detetction via multi-corpus training", 53 (2015) 196–207.

[12] SEGURA-BEDMAR et al.," Exploring spanish health social media for detecting drug effects", BMC MED. INFORM. DECIS".MAKING 15 (SUPPL. 2) (2015) S6

[13] CHEE et al., "Predicting adverse drug events from personal health messages", AMIA ANNUAL SYMPOSIUM PROCEEDINGS", 2011, AMERICAN MEDICAL INFORMATICS ASSOCIATION, 2011, P. 217.

[14] WU et al.,"Exploiting online discussions to discover unrecognized drug side effects, methods inform", MED. 52 (2) (2013) 152–159.

[15] Nikfarjam, A., & Gonzalez, G. H. (2011). Pattern mining for extraction of mentions of adverse drug reactions from user comments. In AMIA Annual Symposium Proceedings (Vol. 2011, p. 1019). American Medical Informatics Association. K. PORTIER, G.E. GREER, L. ROKACH, N. OFEK, Y. WANG,

[16] P. BIYANI, M. YU, S. BANERJEE, K. ZHAO, P. MITRA, J. YEN," understaning topics and sentiment in an online cancer survivor community ", J. NATL. CANCER INST.—MONOGR. 47(2013) 195–198.

[17] F. Liu, L.D. Antieau, H. Yu, Toward automated consumer question answering:automatically separating consumer questions from professional questions inthe healthcare domain, J. Biomed. Inform. 44 (6) (2011) 1032–1038

[18] M. ABDUL-MAGEED, M. DIAB, S. KÜBLER, SAMAR, "Subjectivity and sentiment analysis for arabic social media",COMPUT. SPEECH LANG. 28 (2014) 20–37

[19] Asghar MZ1, Ahmad S2, Qasim M1, Zahra SR1, Kundi FM1.SentiHealth: creating health-related sentiment lexicon using hybrid approach. 10.1186/s40064-016-2809-x 2016.

[20] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, G. Gonzalez, Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks, in: Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, Association for Computational Linguistics, 2010, pp. 117–125.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Student Mobility in North African Countries and the Need for an Information System for its Management

Majida Laaziri[1], Khaoula Benmoussa[1], Samira Khoulji[1] and Kerkeb Mohamed Larbi[2]
[1,2]Information System Engineering Resarch Group
[1]National School of Applied Sciences of Tetouan.
[2]Faculty of Sciences of Tetouan. Abdelmalek Essaâdi University
majida.laaziri@gmail.com, Benmoussakhaoula24@gmail.com, khouljisamira@gmail.com, kerkebml@gmail.com

ABSTRACT

University cooperation is a major concern in North Africa as in the rest of the world. Indeed, the cooperation of the universities of these countries with their counterparts abroad is very active. The number of actions identified for the countries of North Africa is higher than that recorded for the countries of sub-Saharan Africa as a whole. This means that these countries have thirsted for openness to the world and knowledge. Several partnership programs fund cooperation actions with higher education institutions in North Africa and encourage student mobility, which is a source Enrichment for the education system. In North Africa, the mobility of students, researchers and staff constitutes a great wealth for the education system and one of the essential dimensions of the internationalization of higher education.

Student mobility is of great importance at the heart of the missions and strategic plans of universities, as an essential criterion to take into account from the point of view of the evolution of the university system. It is seen as a policy instrument that globally links higher education systems and promotes the flow of talent and the development of a highly skilled workforce. Mobility presents crucial issues not only in the academic Also in the education of citizens open to the world and to society.

As a result, international mobility is a linguistically and culturally essential experience for students, but it is a discontinuity in their academic curriculum. International mobility is favored by states; it is a wealth both for countries of departure and destination

This article analyses the incoming and outgoing mobility of foreign students from and to the countries of North Africa. This mobility is massive; its rate is notoriously one of the highest in the world since the late 1980s. It involves interactions between students and researchers at international level and it is also a vehicle for knowledge exchange and a criterion for developing the potential for research and innovation, and Improvement of the performance of the higher education system in North Africa.

This article also provides the basis for a cooperative information system capable of managing the inbound and outbound student mobility flows of these countries.

DOI: 10.14738/tmlai.54.3428
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3428

1 Introduction

The concept of the internationalization of higher education is constantly evolving. As this has created new responses to the challenges of globalization, and the universities have entered into more complex international relations, the concept has ceased to refer only to international education in order to integrate the international dimension into curricula And institutional culture.

Internationalization promotes the cooperation and integration of universities with their peers abroad; these links of academic cooperation make it possible to establish relations and networks, exchange experiences and good practices, and Generation of ideas and knowledge. It facilitates the transfer of technology, innovation and the strengthening of scientific capacities in countries[1-3].

From exchange comes enrichment. A mutual enrichment promotes excellence through student mobility, technological transfer, innovation transfer, the pooling of good practices and the mobility of teachers, researchers and staff.

In North Africa and elsewhere in the world, the mobility of students, researchers and staff constitutes a great wealth for the education system and one of the essential dimensions of the internationalization of higher education.

2 Study abroad

In addition to the skills specific to the professional field envisaged, the experience of training abroad is a major cultural asset for university students, and makes the difference in any approach to professional integration.

Participation in a program of studies or internships abroad is an appreciable opportunity for encounters, a factor of maturity, development of personal qualities of openness, language development, autonomy, and on the academic level, a new method of learning. Expatriation, even temporary, can be considered as a career accelerator.

There are several starting options for a student wishing to perform an international mobility during his studies. The various ways of exploring the world vary from one university to another; To date, students enrolled in a university program may leave their university either through an inter-university exchange, internship, language development, practical or field course, A dissertation or a thesis, a research or a completely autonomous way to pursue a program offered abroad.

3 Student mobility

Student mobility is of great importance at the heart of the missions and strategic plans of universities, as an essential criterion to take into account from the point of view of the evolution of the university system. It is seen as a policy instrument that globally links higher education systems and promotes the flow of talent and the development of a highly skilled workforce. Mobility presents crucial issues not only in the academic Also in the education of citizens open to the world and to society.

As a result, international mobility is a linguistically and culturally essential experience for students, but it is a discontinuity in their academic curriculum. International mobility is favored by states; it is a wealth both for countries of departure and destination[4-6].

4 Motilities Cooperation projects for the countries of North Africa

There is today a real need for education, training, mobility and integration of young people in the countries of North Africa. Youth in these countries is confronted with the difficulties of professional integration. The youth unemployment rate is at an alarming level.

For this reason, it is now necessary to strengthen the academic cooperation between the countries of North Africa and other international countries in order to promote student mobility at all levels of education. These co-operative links are aimed at fostering the improvement and modernization of higher education systems, as well as vocational training systems through exchanges of good practice, mobility and cultural exchanges.

These cooperative links with other international countries are essentially in a multilateral framework with the support of the European Union, which finance programs and projects for cooperation with the countries of North Africa that promote student mobility. Such as Erasmus Mundus, the OMJ, Tempus projects 2008-2013, etc. [7-9].

These cooperative programs offer scholarships for students pursuing their studies abroad, aim to increase mobility and cooperation, promote the development of higher education systems, but also facilitate the process of economic and social reform, And development in partner countries.

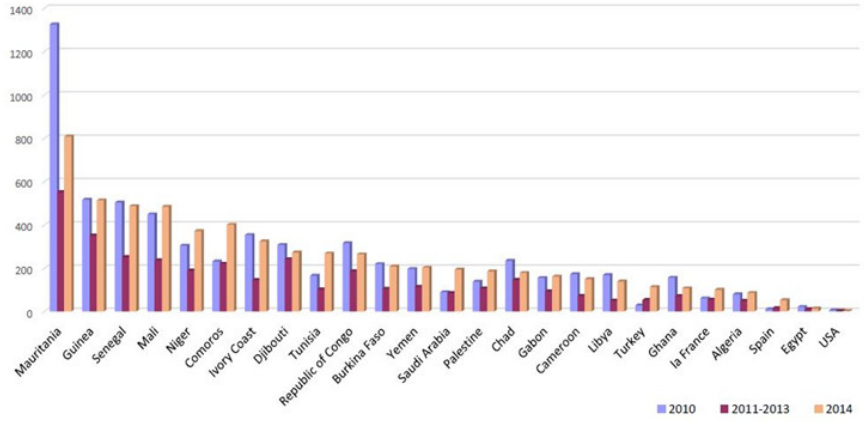


Figure 1 : Incoming Student Mobility in Morocco

5 Incoming and Outgoing student mobility in the countries of North Africa: case of Morocco, Tunisia, and Algeria.

5.1 Incoming and outgoing student mobility in Morocco

5.1.1 Incoming student mobility

In terms of Incoming mobility of foreign students, Morocco has been a country of student expatriation for several years. This mobility is massive, its rate is notoriously one of the highest in the world since the late 1980s, it has become a highly coveted country by foreign students from almost all walks of life[10-11].

On the side of international students hosted, and according to UNESCO statistics, Most of the mobility students in Morocco come from sub-Saharan Africa: Mauritania, Guinea, Senegal, Mali and Nigeria. These students benefit from the cooperation agreements signed with Morocco, offering scholarships to African students. These agreements represent the main part of the Moroccan-African cooperation whose

objective is to promote economic and commercial relations between Morocco and its African space and to improve its political position on the continent.

5.1.2 Outgoing student mobility

Outgoing student mobility is necessary for scientific excellence, which involves interactions between students and researchers at international level, it is also a vehicle for knowledge exchange and a criterion for developing the potential for research and innovation, and Improvement of the performance of the higher education system in Morocco.

According to statistics from the UNESCO database, Morocco ranks among the top countries of origin of mobile students in the Mediterranean region, and ranks third behind France and Turkey. France is by far the destination of the Mediterranean area most appreciated by Moroccan students in front of Spain, Germany, Ukraine, Italy and Canada. These countries offer scholarships to young university graduates, teachers and researchers in all branches of science, as well as in the fields of music and the visual arts.

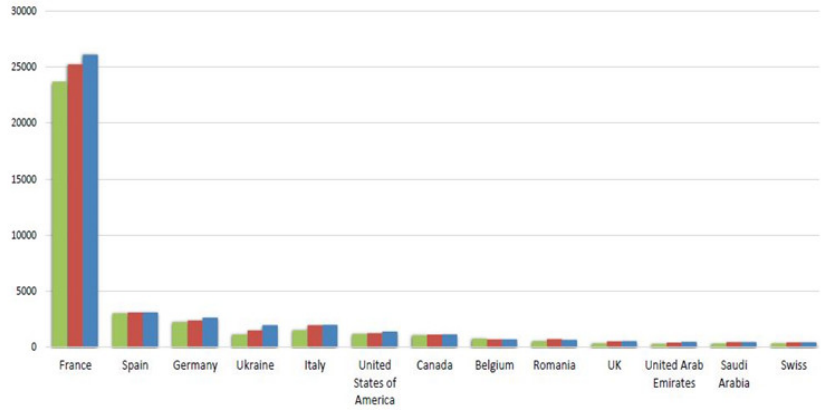


Figure 2 : Outgoing Student Mobility of Morocco

5.2 Incoming and outgoing student mobility in Tunisia

5.2.1 Incoming student mobility

In recent years, the Tunisian higher education system has achieved a remarkable qualitative leap, thanks to the reforms undertaken which have made it possible to equip all the regions with university institutions and to create a network of universities developing scientific research components. Moreover, the teaching is free and mainly taught in French, which attracts foreign students, especially from francophone Africa. Finally, the post-revolutionary government has developed a policy of attractiveness to the countries of sub-Saharan Africa, which has increased the number of foreign students continuing their studies in Tunisia. Most of these students come from Congo, Libya, Mali, Cameroon, Mauritania, Côte d'Ivoire and Maghreb countries such as Morocco[10-11].

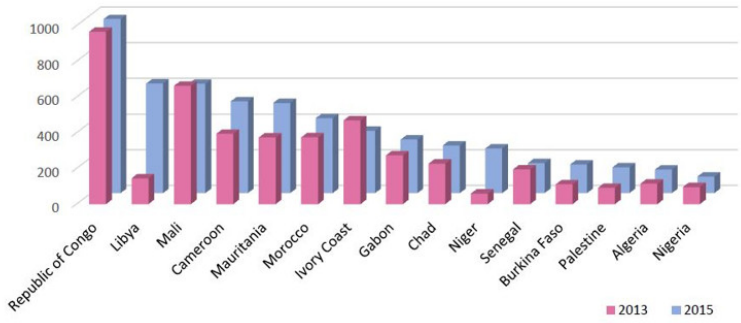


Figure 3: Outgoing Student Mobility of Morocco

5.2.2 Outgoing student mobility

The presence of foreign students in universities abroad is a source of enrichment and cultural creativity in Tunisia. Most Tunisian students are still attracted mostly by France, thanks to the use of French and the Tunisian higher system, which is strongly inspired by the French system, which greatly facilitates the temporary mobility of Tunisian students to France, as well as The setting up of partnerships and double degrees. After France comes Germany as the second host country for Tunisian students, using its scholarship and research programs for Tunisian students and researchers. After that, Romania was ranked third, followed by Canada, Italy, United States, Ukraine, Switzerland and Middle-Eastern countries such as Saudi Arabia.

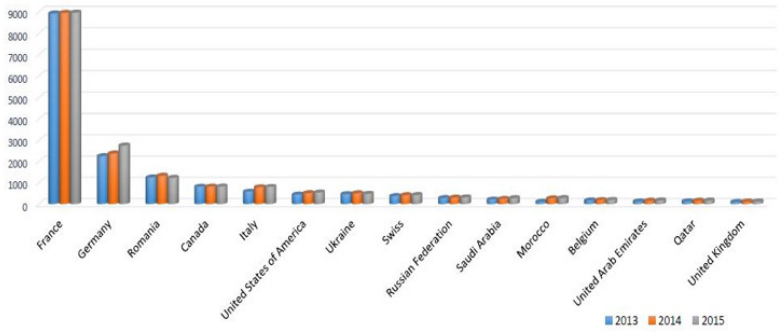


Figure 4 : Incoming Student Mobility in Tunisia

5.3 Incoming and outgoing student mobility in Algeria

5.3.1 Incoming student mobility

For incoming mobility in Algeria, Algeria welcomes an increasing number of international mobility students studying in Algerian higher education institutions according to Unesco statistics. The distribution of these students by country of origin is not available in the UIS statistical database on mobility, but most of these students are probably from sub-Saharan Africa and come from Botswana, Burkina Faso, Burundi, Cameroon, Chad, Congo, Côte d'Ivoire, etc. [10-11].

5.3.2 Outgoingstudentmobility

For the outgoing mobility of Algeria according to Unesco statistics (UIS), it is almost exclusively geared towards France, which receives almost 90% of the total number of Algerian students in mobility. Outside

France, the 2nd Position is for Canada, thanks to its good bilateral relations with Algeria, and the common use of the French language representing a factor of rapprochement between the two countries. In the 3rd position comes Saudi Arabia and the United Arab Emirates with their economic relations, followed by Malaysia, Belgium, Spain which has put in place agreements to develop trade with Algeria. The United States of America, Italy, ...

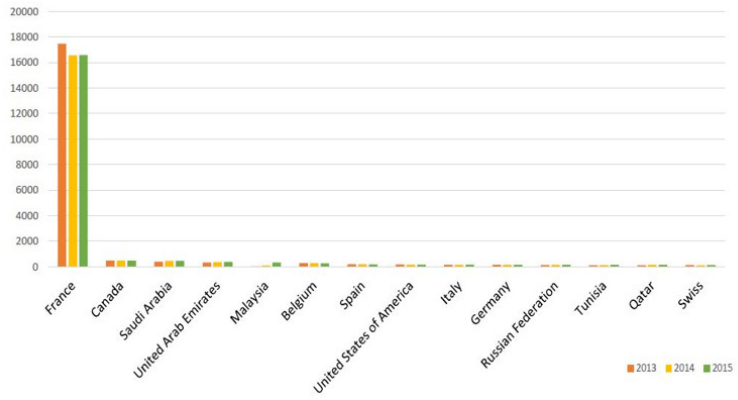


Figure 5: Outgoing Student Mobility of Algeria

In order for a university or country to achieve high levels of performance in terms of student mobility, it needs to improve the efficiency of their cooperation and partnership actions. An information system is a tool that can be used in order to achieve high levels of efficiency in monitoring and evaluation of this kind of actions.

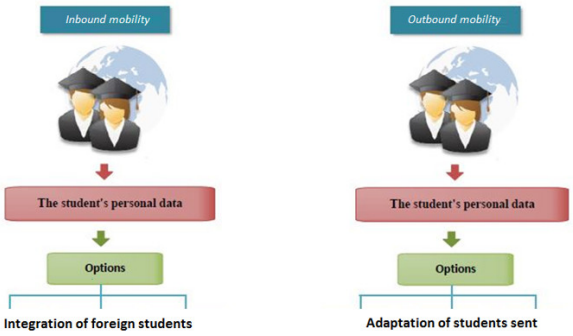


Figure 6: inbound or outbound mobility

Information systems can be used to create new services aux administrators staff and also an entirely new cooperation model. Such model can describes how an university produces, delivers and how they sell a product or service to create wealth.

Such a model describes how a university establishes co-operative relationships with other institutions of higher education, how it sends and receives students, and how it monitors these students in order to to succeed the decided mobility.

We developed an information system capable of providing real-time data on the situation of students who benefited from inbound or outbound mobility. This system was designed in consultation with all the players involved in student mobility in North African countries.

6 Conclusion

In the context of the globalization of academic competition, international exchanges are of a strategic nature for research and higher education in Europe and in North Africa.

University mobility is now a must. Creator of culture and open-mindedness, a key element of academic success and of professional integration, it is a tremendous lever for the recognition and international influence of universities, Position in the global race to excellence ... and also to make it known.

REFERENCES

[1] Laure Endrizzi, " La mobilité étudiante, entre mythe et réalité",2010.

[2] Brian D. Denman, « La Gestalt revisitée : retombées et évaluation de la coopération universitaire internationale », Politiques et gestion de l'enseignement supérieur, 1/2004 (no16), p. 73-92.

[3] Eric Gobe, Catherine Marry, Fran¸coiseChamozzi. « Mobilités internationales et attaches familiales des diplômes maghrébins de grandes écoles ». Sociologies, Toulouse : Association internationale des sociologues de langue française, 2013, pp.1-15.

[4] Abdelwahab Hafaiedh, « La mobilité internationale des étudiants tunisiens : Aide publique et mobilité institutionnelle»,2007.

[5] Nathalie Havet, « Mobilité internationale des étudiants du supérieur et débuts de vie active », 2013 Catherine Agulhon, Ridha Ennafaa, « Les étudiants étrangers. Des trajectoires spécifiques ? ». Les vies étudiantes. Tendances et inégalités, La Documentation française, 2016, 978-2-11-010267-6.

[6] Sylvie Mazzella. « Introduction. Une "libéralisation d'Etat" de l'enseignement supérieur ?: Mutation internationale et évolution maghrébine »,Maisonneuve et Larose. L'enseignement supérieur dans la mondialisation libérale, Karthala, pp.356, 2008, IRMC, 978-2-7068-1990-2.

[7] Terrier Eugénie, « Les migrations internationales pour études : facteurs de mobilité et inégalités Nord-Sud », L'Information géographique, 4/2009 (Vol. 73), p. 69-75.

[8] Souley Mahamadou Laouali et Jean-Baptiste Meyer, « Le Maroc, pays d'accueil d'étudiants étrangers », Hommes et migrations, 1300 | 2012, 114-123.

[9] Livre blanc, « Coopération Universitaire pour Le Développement entre l'Afrique et l'Europe : Relever Les défis régionaux et globaux ».

[10] Unesco Institute of Statistics (UIS) [Online] Available :http://data.uis.unesco.org

[11] International student mobility [Online] Available:http://www.uis.unesco.org

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Comparative Study of Governance Information Systems for Scientific Research

Khaoula Benmoussa[1], Majida Laaziri[1], Samira Khoulji[1] and Kerkeb Mohamed Larbi[2]
[1,2]Information System Engineering Resarch Group
[1]National School of Applied Sciences of Tetouan.
[2]Faculty of Sciences of Tetouan.
Abdelmalek Essaâdi University
Benmoussakhaoula24@gmail.com, majida.laaziri@gmail.com, khouljisamira@gmail.com, kerkebml@gmail.com

ABSTRACT

The research and innovation effort is a major asset in international economic competition. Research and technological development are key areas to achieve this, contributing to economic growth and job creation. In order to achieve the objective of becoming the most competitive and dynamic knowledge-based economy in the world, several government have developed in recent years information systems to manage their scientific research.

As part of the development of instruments to support the strategic management of research at an institution, in particular a university, the design of an information system that meets the requirements of a high level of operational, tactical and strategic institutional authorities seems paramount. International models of research governance or simply the remarks of faculty-researchers at universities help to build a performance management information system.

In this article we will present a comparative study between the information systems of scientific research, in particular the Chile Research Management System (SGI) and the Argentine Science and Technology Information System (SICyTAR), Scientific Information System of Andalusia Spain (SICA), GRAAL of France (management of research data, application of laboratory activities) and others.

1 Introduction

The governance of scientific research reveals the roles of different actors in the research system, how the rules of the organization of the system work, how decisions are taken and how changes in the overall research system come into being

This international comparative study on the governance of innovation was launchedin order to see whether North African countries could learn from governance practices in other countries. North African universities are particularly interested inresearch governance as key actors are rethinking how the research governance system could be improved to make research policy more effective.

DOI: 10.14738/tmlai.54.3426
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3426

The improvement of a scientific research information system on its functionalities, structure, strategies and techniques, etc. requires a comparative study of the existing research information systems and needs also to take into account the remarks of academic researchers at universities.

This paper is based on a comparative study in six countries: France, Spain, Italy, Argentina, Chileand Morocco. The objective of the studyis to describe national models of public governance and examples of good practice.

2 Research Management System (SGI), Chile

Since 2001, the decision to transform the University of Talca into a complex university was adopted, which made it actively involved in the development of research and graduate programs. This decision results in higher academic productivity, particularly in research. Since then, a steady increase has been observed in the number of projects spent in various competitive funds, the number of indexed publications, and enrollment in graduate programs.

The University of Talca has developed the Research Management System (SGI) to support the academic activities of their researchers. The incorporation of SGI at the University of Talca has contributed to a significant increase in the number of active researchers and the number of research projects carried out, the number of publications and the allocation of external and internal funds.

Initially, the system was structured to include a range of information services for the exclusive use of specific users according to their profile, researchers, visitors and others in terms of programs, projects, events and products resulting from the research activities developed. The strategic level of the system aims to have a set of indicators and related statistics of available research capabilities[1-3].

Figure 1: The functional perimeter of SGI

3 Information System of Science and Technology (SICyTAR), Argentina

The Ministry of Science, Technology and Productive Innovation of Argentina has set up the Science and Technology Information System (SICyTAR) to facilitate and unify access to information on Scientists, technicians and their jobs. SICyTARwas created in 2002. Meanwhile, the strategic importance of SICyTARis to keep a unified register constantly updated; produce detailed statistical information to develop

indicators and to evaluate science and technology policies; provide a useful tool for networking among researchers as well as with government and business sectors.

With this important inter-agency support, the Ministry has faced the development of CV registry according to the agreed design: CVar. The CVar is a central application loaded with school data, the result of joint work between the Ministry and the National Council for Scientific and Technical Research, and it is part of SICyTAR.

This new platform also includes an InfoSICYTAR service, with general consultation statistics that presents two platforms (GEICyT, SICyTARBox) for exclusive use by SNCTI institutions.

Among the new features integrated in SICyTAR is BuscaCV.ar : A thematic research system of people that promotes the exchange of information between individuals, groups and institutions. This tool allows accessingasummarized curriculum vitae data of each teacher and researcher [4].

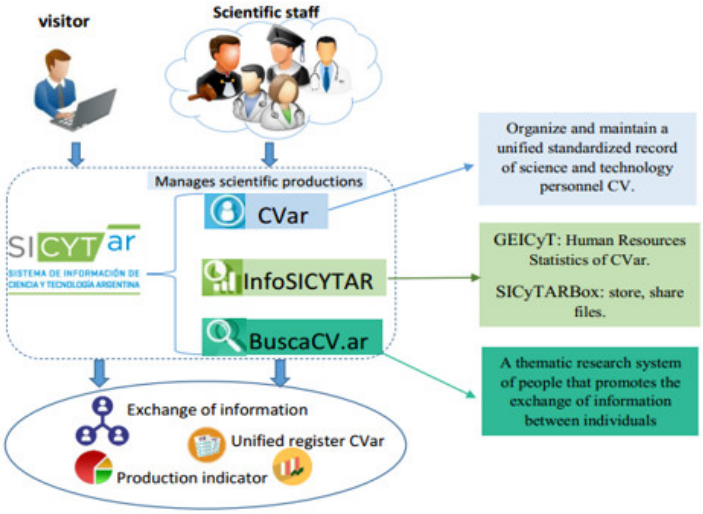


Figure 2: The functional perimeter of SICYTar

4 Scientific Information System of Andalusia (SICA), Spain

S.I.C.A was born as a research project in July 2001. Thanks to a cooperation agreement between the Ministry of Education and Culture of the Government of Andalusia and the University of Granada in order to meet the specific needs of the operational management organisms to promote research and technological development in Spain.

the Scientific Information System of Andalusia, SICA, can be defined as a set of people, processes and equipment designed, constructed, operated and maintained to collect, record, process, store, retrieve and display information related to activities and results produced by researchers in their development centers or in collaboration with other national or international institutions

All these factors have databases that store information to producers and agents (researchers, scientific institutions, research groups, R & D, etc.) on the one hand, and the results of their research activity (Transfer of scientific production technology, curriculum teaching and research, etc.)on the other hand. These databases are usually ad-hoc systems or bibliographic data[5-6].

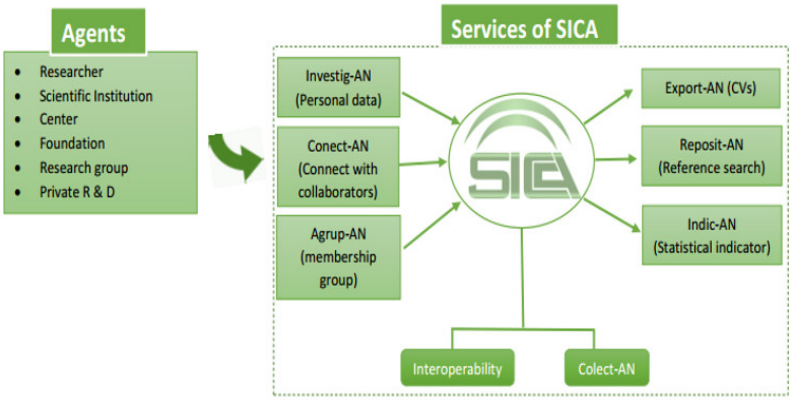


Figure 3: The functional perimeter of SICA

5 Management of Research Data, Application of Laboratory Activities (GRAAL), France

GRAAL (Management of Research Data, Application of Laboratory Activities) is a software program launched in 2000 by the Inter-University Center of Computing in Grenoble. Its piloting was entrusted to a Group of scientific interest (GIS) gathering interested public partners. At the end of 2009, GRAAL was deployed in some fifty establishments.

GRAAL is an application that aimed to present the research units within the universityin a coherent way, including the personnel and their scientific activities (publications, patents, eventsand equipments) as well as the monitoring of financial resources (Endowments Quadrennial contracts and tenders) and international activities (Conventions of partnership, congresses). It allows the data to be historicized and thus to monitor and manage the transformations of research units, financial and human resources, and indicators. It allows three levels of hierarchy in the institution to enter or access certain data: the research unit, The Research Directorate and the presidency.

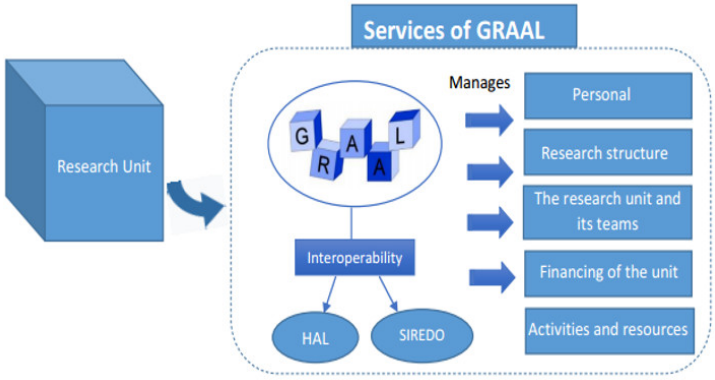


Figure 4: The functional perimeter of GRAAL

6 Moroccan information system for scientific research (SIMArech)

The Moroccan Ministry of Higher Education and Scientific Research has shown great interest in the development of an Information System for Scientific Research. SIMArech was created at Abdelmalek Essaâdi University in 2008 in order to contribute to enhancing the potential of research by highlighting

Figure 4: The functional perimeter of GRAAL

scientific production and Know-how of researchers. SIMArechcould be also used as an instrument of evaluation and could help making decisions.

SIMarech provides a description of the existing situation and a study of the needs for a better optimization of human and material resources. It also allows in a national context of evaluation of the research to engage in a quantitative assessment process that would provide objective criteria for self-assessment and external peer review.It is configurable and allows a transcendent role assignment: teacher-researcher, responsible for the research unit, high education institution and university administrator. Its reporting system makes it possible to print documents, graphs and statistics on various sections specific to research: indexed publications, communications, projects, patents, etc. [10].

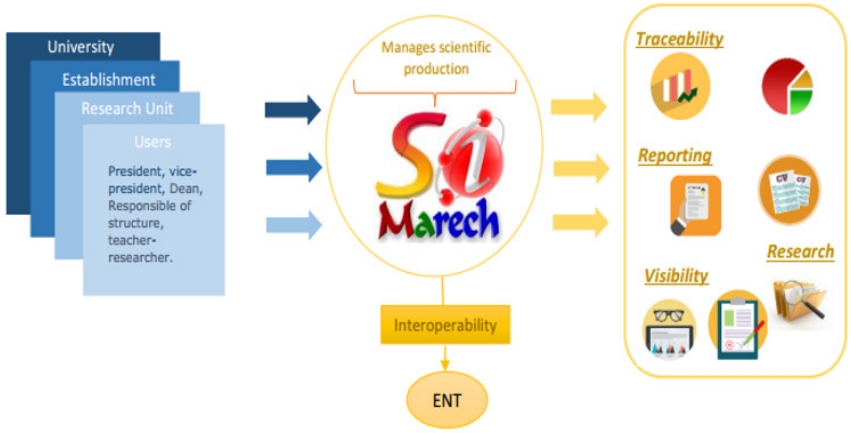


Figure 5: The functional perimeter of SIMArech

7 IRIS (Institutional Research Information System) Italy

IRIS (Institutional Research Information System) is an IT platform for the management of data on research activities (people, projects, publications) adopted by the University of L'Aquila, Italy in 2015. IRIS aims at collecting, managing and storing data related to the University's scientific research output. Professors, researchers, research fellows and Ph.D. students, together with administrative staff members can access the platform using their login credentials in order to catalogue their published research works. IRIS also aims at monitoring research results and enhancing visibility of the University's output. IRIS has five different modules integrated together through standard protocols and interfaces. One of the main components of IRIS is DSpace-CRIS, an open source solution that can also be used as a standalone system, and as such is now installed in seven institutions outside Italy [7].

The system complies with the high standards of representation of scientific entities (people, equipment, installations, projects, etc.) recognized internationally by facilitating access to international scientific networks for Italian universities. The application also allows managing the internal search establishment processes and collect useful data to fulfill several ministerial or European Unionrequirements.

The new system is a valuable aid:

- For government agencies that will monitor the progress of activities, outputs (e.g. products and publications) achievements, and a plan for investment of new funds;

- For researchers and relevant offices, they can obtain an overview of the results of existing research projects, CVs of researchers, bibliographies, and tools available; thus, promoting contacts and collaborations between researchers and the reality of companies.

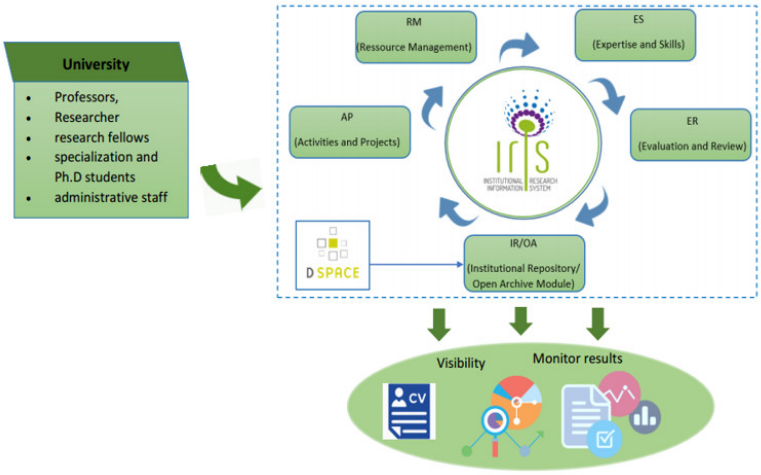


Figure 6: The functional perimeter of IRIS

8 Infocentre research (Infocentrerecherche), France

In 2004 the university paulcézanne Aix-Marseille III launched a project called Infocentre research.The latter is an information system and tool to aid the piloting of the research of an establishment based on the use of a CMS (Content Management System) of the collaborative extranet type and it is totally configurable "in real time". Its flexibility makes it able to adapt dynamically to the specificities of the institution or research unit considered. Thus, this InfocentreResearch is an instrument capable of continuously drawing up a "state of the art" or "dashboard" for these structures.

Emphasis is placed on the fact that local research structures, as well as national structures, have less need of evaluation than of valorization of research. It allows research structures to collect structure data in international databases, add additional metadata, data validation by the relevant actors, the setting up of various reading grids, the composition of dashboards, Etc.

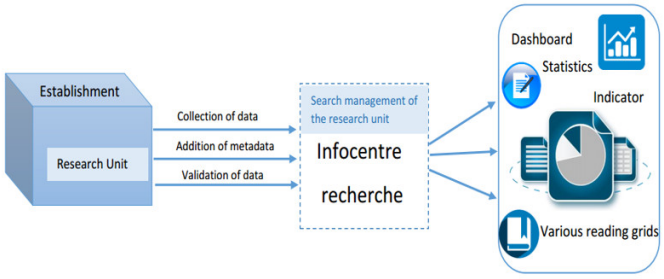


Figure 7: The functional perimeter of infocentrerecherche

9 Integrated Management and Evaluation System SIGEVA, Argentina

Integrated management system and Evaluation- SIGEVA- is a set of computer applications that can be accessed safely through a web platform (intranet) and via an Internet browser. The system was created

in 2005, developed by the department of computer systems organization and CONICET management to generate information, standardize processes, manage and evaluate individuals and research groups, and systematizing information.

This system manages processes, both administrative and academic. It is composed of different modules, module of management of the calls (Management and control, evaluation, decision), Compile module (avoids the duplication of the records), interoperability module, activities module (100% CVarcompatible), Personal Data, Training, Scientific Production, Production Technology, Funds, handles ROLES allowing different users to access certain information and functionality throughout the process[8-9].

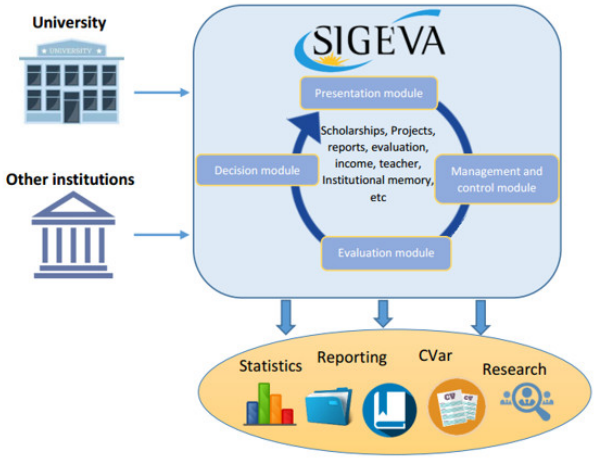


Figure 8: The functional perimeter of SIGEVA

10 Conclusion

In all developed countries, scientific and technological development has become particularly important. Meanwhile, to meet the challenge of being effective in managing research at universities, the amount spent on research and the expectations they deliver to Quality are enormous. Rather than leaving research management, evaluation, impact and reputation to random execution, the institution is better servedwith a research management solution through an information system that responds to the requirements of a high level of operational, tactical and strategic institutional authorities, its researchers, the various stakeholders in research and its results, including business and government.

In fact, the study found that many countries are struggling with similar governance issues. A critical one is how to arrive to a coherent management of science, technology and innovation policy. The study identified that in spite of an existing global research, the need for greater coherence and integration of how are research performers (universities and public research organizations) held accountable for their activities, and how do national and regional actors coordinate their activities.

We have identified various responses to the above questions in the benchmark countries. In all governance systems,there are linear modes of operation function in parallel with more systemic and integrative modes of operation.Partly this can be explained by path dependency and a long cultural heritage, to some extent since some elements of the linear model still fulfill a function in a systemic governance model.

Figure 8: The functional perimeter of SIGEVA

Countries are converging on the idea that there should be a very high-level research policy function to generate a high quantitative and qualitative scientific production.Concerning the synergy between national and regional research activities, we found that in most studied countries, the co-ordination does not exist.

Overall, the governance models in the benchmark countries can be seen as a continuous learning process of adopting organizations and practices to external and internal challenges. The snapshot view that has been provided by this study cannot capture this national specific historical process in great depth. Similarly, good practices cannot simply be transferred from one country to another. Nevertheless, the illustrations how the benchmark countries have tackled some of the key research governance issues, provide food for thought to improve the efficiency of any new research information system.

REFERENCES

[1] R.Claudia, N.José, S.Samaly, S.Cabrera Francisco, Fernández Juan Antonio, Chaichio Juan Antonio, 'Sistema de Información Científica de Andalucía: una herramienta para la evaluación y gestión de los resultados de la actividad científica',31 may.2006. https://core.ac.uk/download/pdf/11880141.pdf

[2] l.manderieux, 'guía práctica para la creación y la gestiónde oficinas de transferencia de tecnología en universidades y centros de investigación de américa latina', June 2011.

[3] F. Palomo, Carlos G. Veloso, y Rodolfo F. Schmal, 'Sistema de Gestión de la Investigación en la Universidad de Talca, Chile', InformationTechnology ,Vol. 18 N1 – 2007, pp. 97-106.

[4] J. Ferrer, C. Clemenza, 'Gestión de la investigación universitaria: Un paradigma no concluido', Multiciencias, vol. 6, num. 2, mayAugust, 2006, pp. 188-193.

[5] F.Manuel Solís Cabrera, 'El sistema de información científica de Andalucía, una experiencia pionera en España',2006.

[6] Scientific Information System of Andalusia (SICA) [online] Available:https://sica2.cica.es

[7] A.Bollini, M.Mennielli, S.Mornati, DT.Palmer,' IRIS: Supporting & Managing the Research Life-cycle', Universal Journal of Educational Research, vol. 4 n. 4, 2016, p. 738 – 743.

[8] Integrated Management and Evaluation System (SIGEVA) [online] Available: http://sigeva.conicet.gov.ar

[9] [9] Management System for Research (SGI) [online] Available: http://sgi.utalca.cl/directorio_new.php

[10] Moroccan Information System for Scientific Research (SIMArech) [online] Available:http://www.sirech.uae.ma/portail/public/index.php

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

SIMARECH 3: A New Application for the Governance of Scientific Research

Khaoula Benmoussa[1], Majida Laaziri[1], Samira Khoulji[1] and Kerkeb Mohamed Larbi[2]
[1,2]Information System Engineering Resarch Group
[1]National School of Applied Sciences of Tetouan.
[2]Faculty of Sciences of Tetouan.
Abdelmalek Essaâdi University
Benmoussakhaoula24@gmail.com, majida.laaziri@gmail.com, khouljisamira@gmail.com, kerkebml@gmail.com

ABSTRACT

In the international context of research assessment, many countries appear to move critically towards a quantitative evaluation policy, which would eventually complement or replace the traditional peer review system. However, infometrics studies, based on the running of large international scientific databases, are confused by the complex organization of their research system. Furthermore, these databases prove to be inadequate to study the local research institutions and units. After a description of the current situation and a needed analysis survey, this proposal presents the design and development of an information system conceived as a tool for monitoring the research carried out by a university or another institution. Based on the use of an effective application developed by a team of Abdelmalek Essaâdi University(UAE) dedicated to conduct collaborative extranet, this information system, called SIMArech "Système d'Information de la Recherche Scientifique", is fully customizable in "real-time". It is available in Arabic, English and French and its flexibility enables it to dynamically adjust to the specificities of any institution having research units. Actually, several universities in three countries are implementing this system to improve their management of scientific research. Several examples are given to demonstrate how this tool is able to continuously develop a "state of the art" or a "dashboard" for these structures and their scientific production. The emphasis is put on the fact that local research structures, as well as national structures, needless research evaluation than valorization of their research.

1 Introduction

Higher education has long been a strongly competitive matter. Governments and university alike keenly watch the rankings produced many multitudes of organism;there aremore than 150 national rankings around the world. Such rankings encourage stratification. One of the metrics is the scientific production that accentuates the differentiation between better and worse universities. The North African model is thus quite different from the European or the North American one, which is a lot less selective and more homogeneous [1].

DOI: 10.14738/tmlai.54.3429
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3429

Because of a high globalizationof the scientific research and the growth in international cooperation[2], attention has shifted from national to international rankings.Governments want top-class universities because the modern economy is driven by scientific research and high-qualified human capital.

Both public and private universities that seek research funding from government and foundations must provide evidence of their accomplishments and capacities. These universities must account for their performance as part of professional accountability protocols[3]. Indeed, every university must have a clear and an evidence-based understanding of the institution's performance towards its goals and mission.

This understanding is achieved and maintained through an ongoing evaluation of all the institution functions. Meanwhile, because research is a central function, the university must evaluate its performance. Data on research performance helps to inform strategic decisions about what areas of research to support or build. It also helps the university leaders understand the institution's position relative to global and domestic standards of research production[4].

Every university must know how much research is conducted, and what its impact was, how many articles of university members were published and in which class of journals, etc. In fact, there are many international databases and editors dedicated to counting, measuring, and comparingthe scientific articles produced by members of universities[5]. However, the research results through publications has become enormous and complex, and itshould not be restricted to scientific articles but should include also several other categories such as thesis, communications, books, prizes, research projects, motilities, patents, facilities, equipment and services etc.In fact, it is so complex and specialized that personal knowledge and experience are no longer sufficient tools for understanding trends or for making decisions by universities or governments[6].

Government and universities must be able to make good decision about what research should be supported and what should not, or which research projects and researchers should receive more support than others. For a good piloting of the scientific research, including all its components, an information system becomes an obligatory requisite to manage better investments in scienceto evaluate the research performance and to establish a sound policy for the development of scientific research[7].

2 The information system SIMArech

Based on the use of an effective application developed by a team of the UAE dedicated to conduct collaborative extranet, this information system, called SIMArech, aims to support research administration by a unique technology solution in order to allow a streamlined and single-point access to up-to-date information for researchers, research staff and administrative staff.

SIMArech is indeed the result of a collaborative way of working. This innovative best-of-breed solution has been designed to fulfill the needs of academic andresearch institutions: SIMArechis an IT platform that makes it easy to collect and manage data on research activitiesand outputs within an organization. Researchers, administrators and evaluators are given all thetools needed to monitor research results, enhance visibility and efficiently allocate available resources.

The modular nature of the system and the flexibilityof its data model facilitate processing, organizing and transmitting information in accordancewith international standards. At the same time,SIMArech provides the chanceto easily shape local and national requirements.

The compliance with international standards and databases of the Web of Scienceoffers several advantages toSIMArech: it strengthens relations amongresearch entities (people, resources, activities,publications, etc.); it supports evaluation activitiesand dissemination; it helps the exchange ofinformation among different Research Systems at national and internationallevels.

Modules are natively integrated with the IT infrastructureof the organization, both with legacysystems (such as HR, accounting databases) andexternal systems (regional, national and internationaldatabases) so that Return on Investment– ROI- is always maximized.

SIMArech supports research by delivering and maintaining innovative and integrated systems that simplify administrative processes in order to:

- Reduce administrative burden.
- Streamline and automate research administration.
- Provide authorized individuals and systems with convenient access to timely information regarding proposals, awards, and compliance.
- Improve the efficiency of administrative offices to meet the demands of a growing workload with limited resources.
- Make the pilotage process of the research addressed faster, easier to understand, and to improve the efficiency of surrounding processes.
- Making the process more transparent by reporting on research status, information sharing and easy data integration between university management systems.

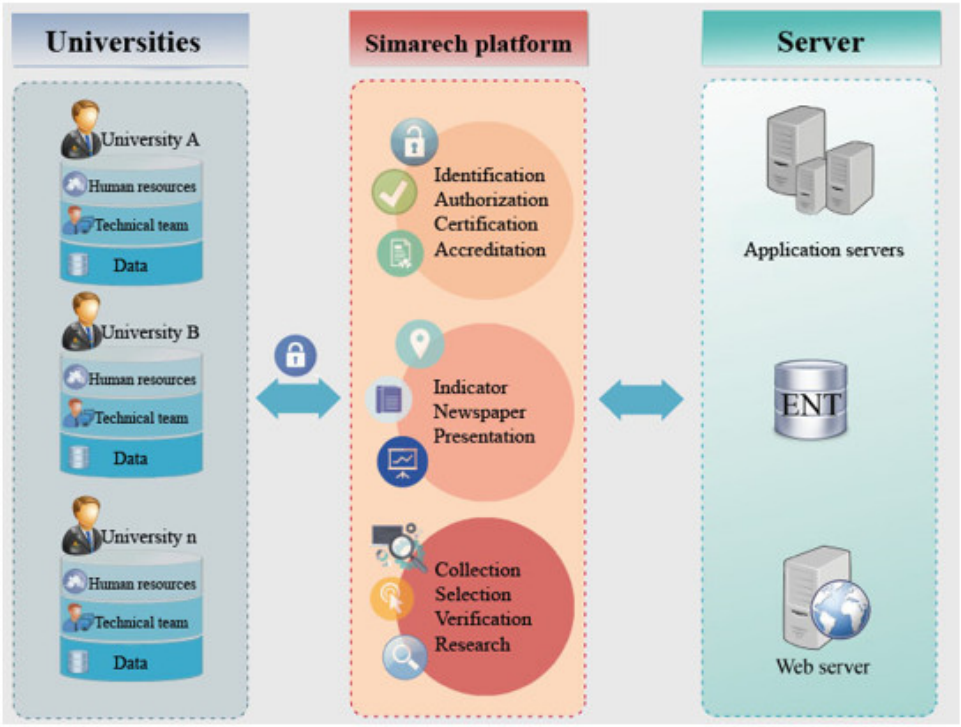


Figure 1: Integration of SIMArech in universities

SIMArech was designed to meet the specific needs of African universities and Governments by:

- Promoting and increasing the visibility, exchanges and communication between scientists and researchers at regional, national or international level.
- Organizing and maintaining unified science and technology registries to produce accurate, reliable and up-to-date statistics in real time.
- Assisting in the overall management of researchers in the research units of the institutions and their scientific results.
- Presenting coherently the research units and their environment.
- Providing up-to-date information on individual research programs.
- Exchanging knowledge at the national and international level, in order to enrich the cultural and professional experience of researchers
- Ensuring interoperability between national scientific and technological databases.
- Establishing a knowledge base that is authorized to submit standardized criteria, and is agreed upon in the evaluation and quality of the results of the scientific activity.

3 The SIMArech Technologies:

SIMArech is based on open-source technologies that the UAE team contributes to develop and maintain. Itwas developed using a Framework called Symfony which was quickly adopted by professionals active in this field following its launch in 2005. This Framework today is a stable environment that is both well-known and recognized internationally. Basically, Symfony is a framework chosen by the community; thus, it can take advantage of all the efforts of the OS developers around it: from automated deployment tools to fully integrated ORMs, from testing frameworks to tutorials and best practices, through native, advanced debugging tools. Symfony is as of today the most complete framework available in the PHP ecosystem when you take in consideration learning curve, integrations, stability and performances. The number of its references attests to this as they have grown significantly since its launch. This Framework is also an active community; developers, integrators, users and other contributors participate in the on-going enrichment of this tool[9-10].

The choice of Symfony as a framework was based on the fact that it is well decoupled, so unit testing becomes very easy since it is possible to mock objects, isolate classes and inject stub dependencies very easily. It provides a first layer for functional testing (with PHPUnit): being an HTTP-centric framework. It provides also a base class that lets simulating HTTP requests and examines the output.

At the end, we can see how Symfony and the ecosystem around it provide the proper toolset to run unit, functional and behavioral tests. From the beginning, we take care of testing and we do not want to waste a huge portion of time doing manual testing. For this reason, we looked for it to take advantage of technology to automate heavy tasks that harms the application development.

This Framework has also a very powerful debug toolbar that would present all these information in order to ease debugging. Database inspection permit to realize how many queries are running and see the SQL of all of them with a nice overview of the time they take, while the profiler itself includes information about every step of the application. Since we are working with a Service-Oriented Architecture, we take advantage of easy integration that Symfony provides for Doctrine[11-12].

Doctrine 2 provides natively support for multiple DB connections and object-relational mappings, we can safely use this tool within Symphony to handle read and writes to different databases without polluting the domain model of each of the services that take advantage of Doctrine. On another note, sharing the data model among different services helps as overcoming some situations where web services or messaging queues are not enough [11-12].

Finally, the probably most powerful concepts of Symphony are the Bundles. They are micro-applications inside the main application, being able to totally separate logics from different domains helps in keeping a clean separation of concerns and autonomously develops every single feature of the domain[11-12].

4 Functional Aspects of SIMArech

The actors of SIMArechcan be definedwithin four roles:

The user or researcher:

- ✓ Dissemination and sharing of knowledge: Participates in increasing the visibility and the diffusion of the scientific knowledge; enriches the system by his/her scientific productions; makes it possible to quote the researchers with whom he/she was collaborating to realize a project...He/she follows and elaborates the activities of a research unit to which he/she belongs: team, laboratory / group, center.
- ✓ Organization within the structure: the researcher must register in a research unit. He/she can apply for membership in a laboratory / group or ask the administration to create a research structure if he/she has formed a group / laboratory, or change his/her membership so he/she will be part of a new group and only this group. "The researcher will be outside the group to which he/she belonged before."

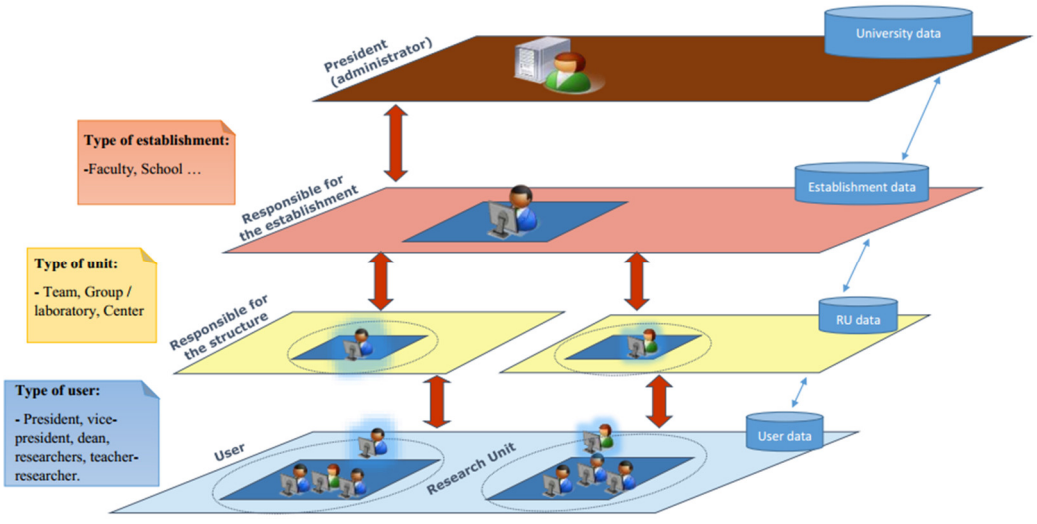


Figure 2: The actors of SIMArech

The responsible of the research unit:

- ✓ Dissemination and sharing of knowledge: participates in increasing the visibility and the diffusion of his/her own scientific knowledge, enriches the system by his/her scientific productions, makes it possible to quote the researchers with whom he/she was collaborating to realize a project ...
- ✓ Team coaching: Records scientific publications, scientific events: congresses, symposiums, round tables, think tanks... for the account of the laboratory and / or research work. Proposes and defines current research work and set priorities: defines a clear search axis. Validates the necessary means (The need for scientific or computer equipment) to carry out a research activity, thus declares a fund offered by the major partner groups or ministries, associations for research, Moroccan fund, to evaluate the progress of his/her team's research.
- ✓ Organization of Research Unit: the research unit can be transformed (URs and teams can explode, merge, regroup, etc ...)

Responsible at the institution level:

- ✓ Dissemination and sharing of knowledge: participates in increasing the visibility and the diffusion of his/her own scientific knowledge, enriches the system by his/her scientific productions, makes it possible to quote the researchers with whom he/she was collaborating to realize a project ...
- ✓ At the establishment level: Manages scientific productions of his/her research units, Evaluates the relevance of the adopted means to develop research within the research units. Develop the institution's budget.

Responsible at the university level:

The responsible at the university level plays the role of an internal animator. Manages all researchers belonging to the university and their scientific productions, controls, verifies, validates and accredits the information and operations...

5 Modeling of the information system (input-output)

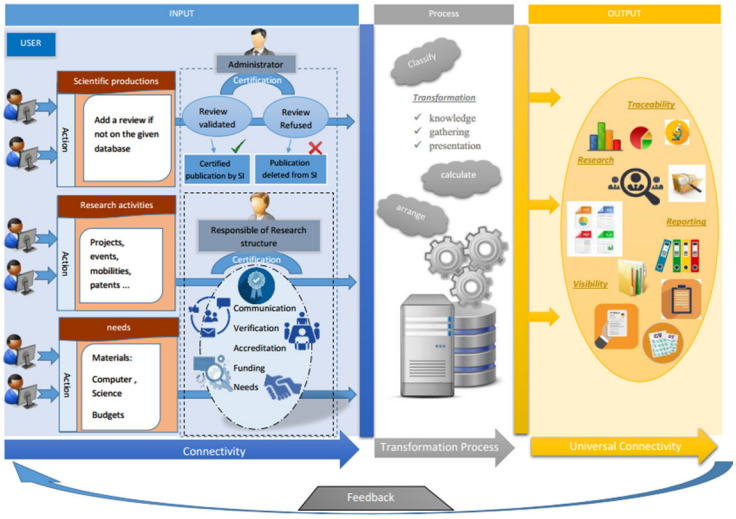


Figure 3 : input output of SIMArech

The SIMArech information system includes inputs and outputs.

At the input level:

All the scientific activities (projects, manifestations, communications, movements, patents ...) entered by researchers, managers of structures, deans, and administrator.

- ✓ The scientific activities entered by the researchers are verified by the responsible of structures.
- ✓ The publications are verified and validated (according to the review) by the administrator.

At the process level:

Consists of a set of processing operations (classify, calculate, organize, etc.) of the data carried out by the SIMArech program.

At the output level:

- ✓ Visibility: CVs, scientific productions and follow-up of the national and international activities (missions, exchanges) the researchers, the research units and the university, as well as the distribution of research units and budgets.
- ✓ Research: research in scientific productions of researchers.
- ✓ Traceability: statistics, dashboards.
- ✓ Reporting: CV, searchingfor structure accreditation file, Reports of scientific productions,
- ✓ Decision: agreements, refusals, Conventions, accreditations, creation of research units,
- ✓ Evaluation: The continuous analysis of scientific productions by those responsible for the structure makes it possible to declare their strengths and their weaknesses at the level of research units, as well as institutions, and universities in general.

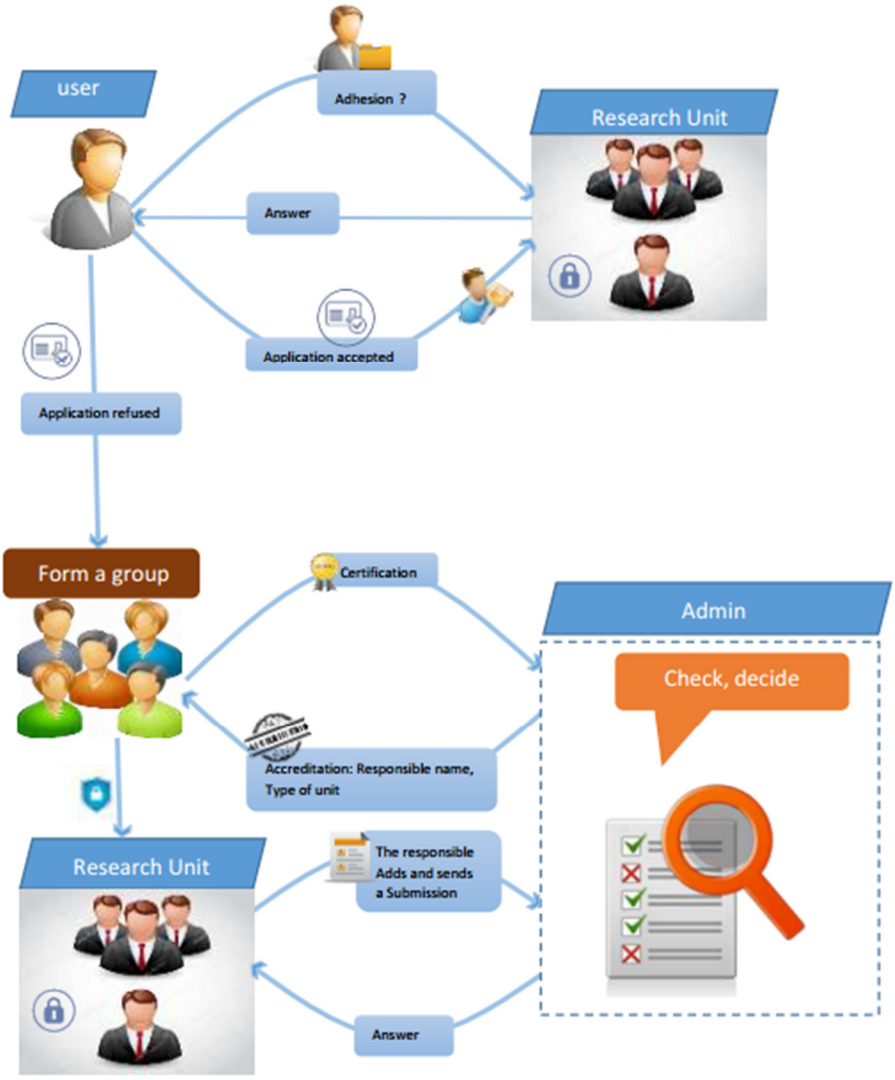


Figure 4: Management of the research unit by SIMArech

6 Conclusion

The scientific production of a university or a higher education institution is highly dynamic given that evolves on a daily basis. It is primordial for decision-makers to have a comprehensive dashboard that can provide them with timely information on research conducted within their institutions.

The SIMArech information system is an excellent collaborative tool that facilitates the research management in a digital way.With advanced menu−driven application tools for research performance assessment and monitoring of university research units, for journal indicators, ranking of universities and mapping, the information system have now reached a stage of high quality, reliable and very informative instruments in research evaluation practice.

REFERENCES

[1] Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S. and van Raan, A.F.J. (2011) Towards a new crown indicator: an empirical analysis. Scientometrics 87, 467–481

[2] VanRaan, A.F.J. (2004) Measuring science. Capita selecta of current main issues. In: Handbook of Quantitative Science and Technology Research (Moed, H.F., Glänzel, W. and Schmoch, U., eds), pp. 19–50, Kluwer Publishers, Dordrecht

[3] Butler, L. and Visser, M.S. (2006) Extending citation analysis to non-source items. Scientometrics 66, 327–343

[4] vanRaan, A.F.J. (2006) Performance-related differences of bibliometric statistical properties of research groups: cumulative advantages and hierarchically layered networks. Journal of the American Society for Information Science and Technology 57, 1919–1935

[5] CABRERA J, DE LA SOTA J, GONZÁLEZ A, GUERRERO H, PRESMANES B, "El papel de los expertos enciencia y tecnología", Revista Madri+d. Monografía 5, p 32-57, 2002

[6] MOED H, Bibliometric indicators reflect publication and management strategies, Scientometrics, V 47, N 2, p. 323-346, 2000

[7] WARTA K, CHARLET V, Enquête 2002 auprès des laboratoires angevins, pour l'Observatoire de la RechercheAngevine, rapport Technopolis France, 2003

[8] ZITT M, BAUIN S, FILLIATREAU G, Production coopérative d'indicateurs inter-institutionnels de politiquescientifique. Indicateurs bibliométriques des institutions publiques de recherche, rapport OST,2002

[9] Natalya Prokofyevaa , Victoria BoltunovaaAnalysis and Practical Application of PHP Frameworks inDevelopment of Web Information SystemsProcedia Computer Science 10451–56 (2017)

[10] 10. SensioLabs. The Symfony Book v. 3.1; 2016. p. 4-68. Available: http://symfony.com/pdf/Symfony_book_3.1.pdf.

[11] 11. PHP Documentation. PHP's memory_get_included_files. Available: http://php.net/manual/en/function.get-included-files.php.

[12] PHP Documentation. PHP's memory_get_usage function. Available: http://php.net/manual/en/function.memory-get-usage.php.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

SIMACoop: a Framework Application for the Governance of University Cooperation

Majida Laaziri, Khaoula Benmoussa, S. Khoulji and ML. Kerkeb
National School of Applied Sciences of Tetouan.
Faculty of Sciences of Tetouan.
Abdelmalek Essaâdi University
majida.laaziri@gmail.com, Benmoussakhaoula24@gmail.com, khouljisamira@gmail.com, kerkebml@gmail.com

ABSTRACT

The management of the cooperation is the process of establishing and ensuring the continuation of constructive and productive relations, consistent with partners. It requires the allocation of time and resources needed to maintain the existing partnership and communicate regularly with partners, and to enable all parties to track the functioning of the partnership. Management system applies to all categories of partners with universities, which include governments, international funders, local authorities, chambers of Commerce and Industry, consulates and embassies, etc. The areas of partnership with universities include academic and research fields but also institutions management to contribute to community service, implement joint projects, provides common services, share and exchange information, implementation joint activities, etc.

For a good piloting of the cooperation, an information system becomes an obligatory requisite to manage better partnerships, to evaluate the research performance and to establish a sound policy for the development of the institution relationship.

Based on the use of an effective application developed by aAbdelmalek Essaâdi University (UAE) team, SIMAcoop is a monitoring system aimed at supporting the governments and universities plan of cooperation, partnership and students exchange programs. The objective of SIMAcoop is to enhance communication, collaboration and integration among universities and their partners, to improve in performance management, strategic planning, and monitoring and evaluation by leveraging the use of spatial data. This emphasis on flexibility to align to national priorities will strengthen mainstream data collection and analysis procedures and, therefore, improve the quality of data reported top management of institutions.

1 Introduction

What any higher education institution can do alone is very limited if not accompanied by partnership relations with external parties who are able to provide support and expertise and advice to the first party. The governments and International funders goal is to support people who will create intellectual property and clusters of high-tech consortium. A great academic and research university is not a sufficient condition for creating such a cluster. It can't do it without cooperating with more great universities and

DOI: 10.14738/tmlai.54.3431
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3431

socioeconomic organism at national and international level. For this reason, all the higher education institution attaches crucial importance to the theme of partnership and the establishment of strategic partnerships that will contribute to facilitate the work of the university and enable it to achieve its mission [1].

The management of the partnership is the process of establishing and ensuring the continuation of constructive and productive relations, consistent with partners. It requires the allocation of time and resources needed to maintain the existing partnership and communicate regularly with partners, and to enable all parties to track the functioning of the partnership.Management system applies to all categories of partners with universities, which include governments, international funders, local authorities, chambers of Commerce and Industry, consulates and embassies, etc. The areas of partnership with universities include academic and research fields but also institutions management to contribute to community service, implement joint projects, provides common services, share and exchange information, implementation joint activities, etc.

For a good piloting of the cooperation, an information system becomes an obligatory requisite to manage better partnerships, to evaluate the research performance and to establish a sound policy for the development of the institution relationship[2].

The needs of universities at this level are very important. Such a system of steering the cooperation should allow to detect the stronger points in the cooperation to overcome the own weaknesses, reduce costs, profit from the exchange of experiences, skills and competencies and raise the level of efficiency and productivity[3]. Thisinformation system must also facilitate the process of managing partnerships signed between the university and between various kinds of partners. It should permit to track the implementation of partnerships signed provisions and the extent to which each party is abided to its obligations and undertakings entrusted to it, and respect for agreements and memorandums of understanding to the missions of the partners and their strategic goals[4].

2 The information system SIMAcoop

Based on the use of an effective application developed by a UAE team, SIMAcoop is a monitoring system aimed at supporting the governments and universities plan of cooperation, partnership and students exchange programs. The objective of SIMAcoop is to enhance communication, collaboration and integration among universities and their partners, to improve in performance management, strategic planning, and monitoring and evaluation by leveraging the use of spatial data. This emphasis on flexibility to align to national priorities will strengthen mainstream data collection and analysis procedures and, therefore, improve the quality of data reported top management of institutions.

SIMAcoop is a coherent global information system with three main components:

- Management of participation in cooperation projects.
- Management of agreements and agreements signed with other national and international universities.
- Managing the mobility of foreign students.

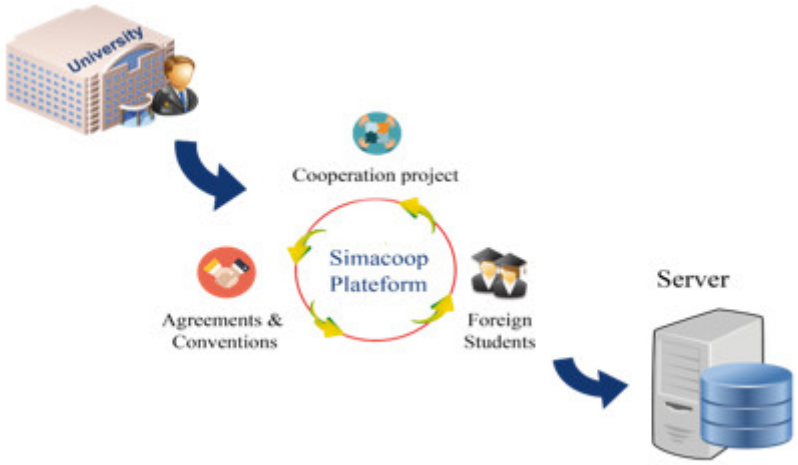


Figure 1: Platform SIMAcoop

- ✓ The cooperation project component manages the general information of the projects, the information concerning the partners who contribute to the implementation of the cooperation projects and programs, the activities, the budget and the expenditure.

- ✓ The agreements and conventions section manages the information concerning the framework agreements and specific agreements signed by the university, the signatories, the actions and the budget of these agreements.

- ✓ The foreign students section deals with the management of information concerning foreign students benefiting from cooperation agreements signed with universities to pursue their higher education.

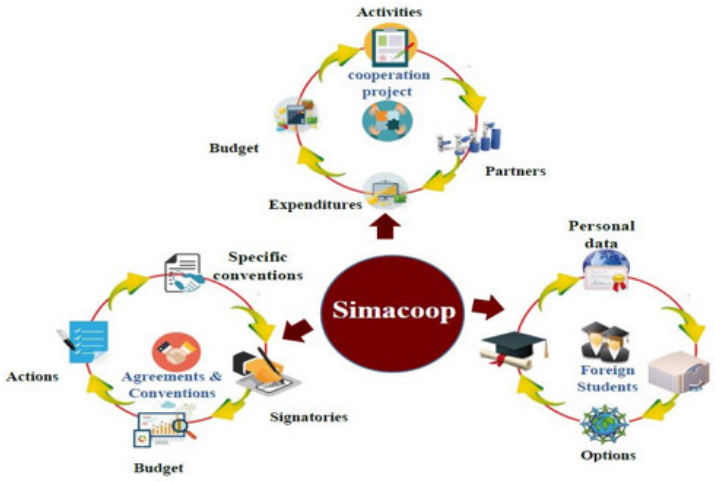


Figure 2: The different parts of the information system SIMACOOP

3 The SIMAcoopTechnologies:

SIMArech is based on open-source technologies that the UAE team contributes to develop and maintain. It was developed using a Framework called Symfony which was quickly adopted by professionals active in this field following its launch in 2005. This Framework today is a stable environment that is both well-known and recognized internationally. Basically, Symfony is a framework chosen by the community, thus it can take advantage of all the efforts of the OS developers around it: from automated deployment tools to fully integrated ORMs, from testing frameworks to tutorials and best practices, through native, advanced debugging tools, Symfony is, as of today the most complete framework available in the PHP ecosystem when you take in consideration learning curve, integrations, stability and performances. The number of its references attests to this, as they have grown significantly since its launch. This Framework is also an active community; developers, integrators, users and other contributors who participate in the on-going enrichment of this tool [5-7].

The choice of Symfony as a framework was based on the fact that it is well decoupled, so unit testing becomes very easy since you can mock objects, isolate classes and inject stub dependencies very easily. It provides a first layer for functional testing (with PHPUnit): being an HTTP-centric framework. It provides also a base class that lets you simulate HTTP requests and examine the output.

At the end, we can see how Symfony and the ecosystem around it provide the proper toolset to run unit, functional and behavioral tests. From the beginning, we take care about testing and we don't want to waste a huge portion of time doing manual testing. For this reason, we looked-for to take advantage of technology to automate heavy tasks that harms the application development.

This Framework have also a very powerful debug toolbar that would present all of these informations in order to ease debugging. Database inspection permit to realize how many queries are running and see the SQL of all of them, with a nice overview of the time they take, while the profiler itself includes informations about every step of the application. Since we are working with a Service-Oriented Architecture, we take advantage of easy integration that Symfony provides for Doctrine [5-7].

Doctrine 2 provides natively support for multiple DB connections and object-relational mappings, we can safely use this tool, within Symfony, to handle read and writes to different databases without polluting the domain model of each of the services that take advantage of Doctrine. On another note, sharing the data model among different services helps as overcoming though situations where webservices or messaging queues are not enough [5-7].

Finally, the probably most powerful concepts of Symfony are the Bundles. They are micro-applications inside your main application, being able to totally separate logics from different domains helps you in keeping a clean separation of concerns and autonomously develop every single feature of your domain[5-7].

Figure 3: Architecture of the actors

4 Functional structure of SIMAcoop actors

In SIMAcoop system, we distinguish mainly three actors who are the following:

- The project manager: he is the project manager who already has an account created in the SIMAcoop platform, he is in charge of project management of cooperation, monitoring of the budget and the project development team, and he has the right to introduce into the system:
  - General information about the cooperation project, partners, activities, budget and expenditure of the project.
  - The general information of the agreements and conventions, the signatories, the budget, and the actions of these conventions.
  - Personal data of foreign students and information on their graduate studies.
- The Establishment administrator: the dean of the institution, responsible for the addition of the project managers belongs to his institution in the SIMAcoop platform, and can be a project manager.
- The university administrator: he is the president of the university, the manager of the platform; he has total visibility on the databases. He is responsible for managing the entire system and responsible for creating and validating accounts so it specifies the access rights of each user, he can be a project manager.

5 Modeling of the information system (input-output):

In the SIMAcoop information system, information management is based on the input-process-output (IPO) model.

- ✓ **Input :** The different users of the system (Project manager, establishment administrator, university administrator) introduce the information and data concerning their cooperation projects, agreements and conventions, and foreign students to the platform SIMAcoop.
- ✓ **Process:** the process of the system receives the data from the users, these data passes a set of operations of transformations, processing, calculation, analysis).
- ✓ **Output :** data processed by the system process is transformed into project reports of cooperation and conventions, budgets, expenditures, CVs, statistics, search engines.

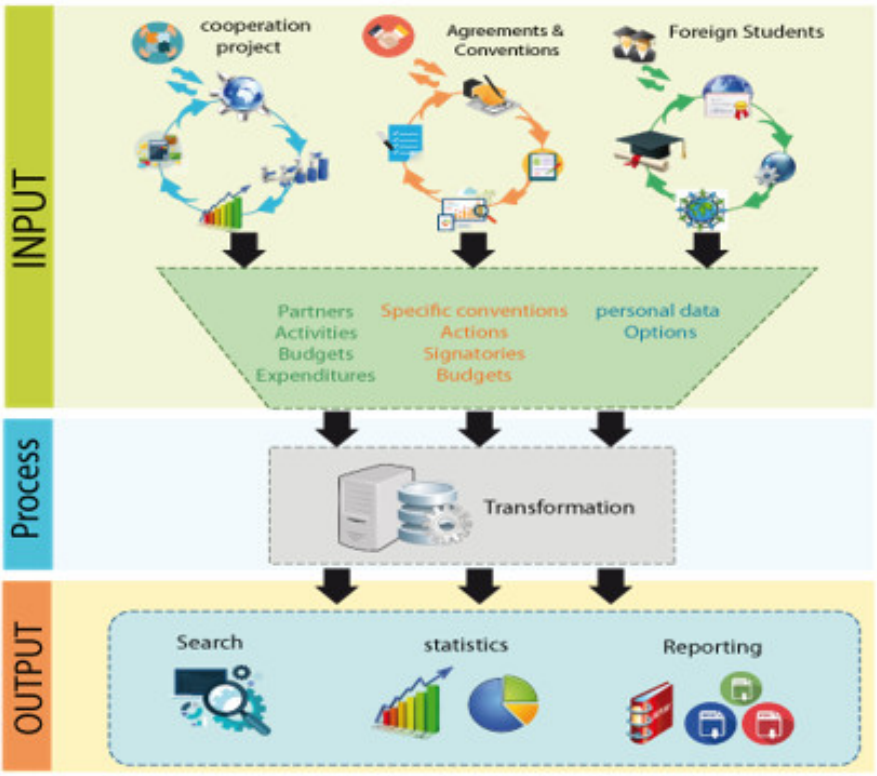


Figure 4: The input–process–output model

6 Internal references in SIMAcoop:

The project manager has the right to add several cooperation projects, partners, activities, budget and expenditure of each project. To differentiate between the data of each project, the Simaccoop information system process is based on the concept of the internal reference of each data relating to the cooperation project.

The internal reference is an assigned identifier for each record of the cooperation project, including the project number followed by the number of the data associated with this project.

Figure 5:the internal reference of cooperation

The same procedure for the internal references of agreements and conventions, and foreign students.

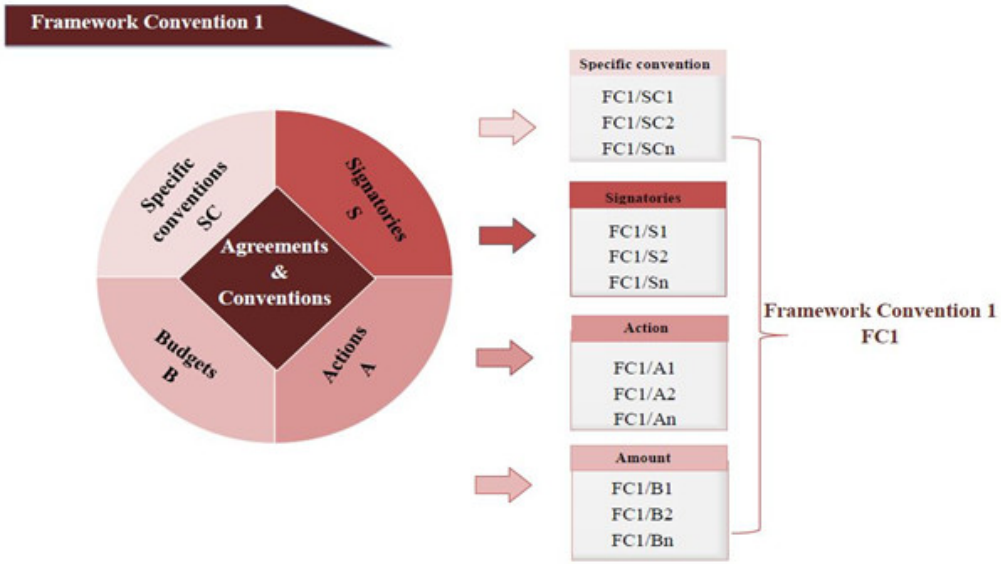


Figure 6: The internal reference of agreement and convention

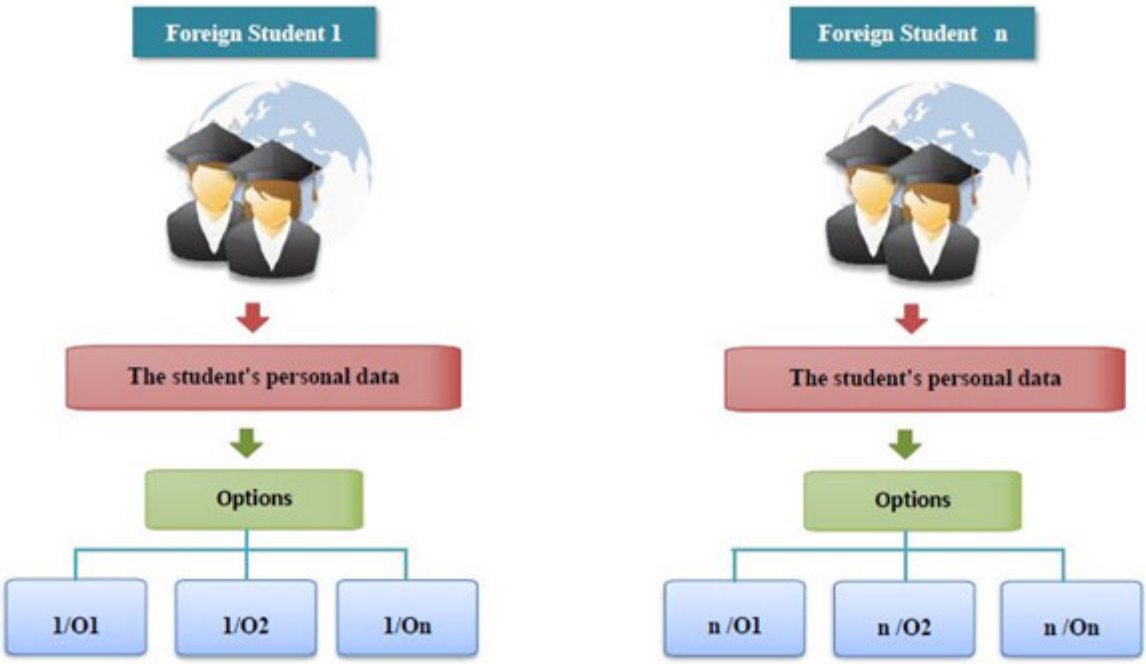


Figure 7:the internal reference of foreign student

7 Warning system in SIMAcoop:

The SIMAcoop information system is enriched by a warning system to send warning notifications to inform the user of the activities of unrealized cooperation projects, framework agreements and specific conventions. Thus, it plays the role of a system of classification.

- For notifications of unrealized activities are based on the comparison of the start date with the system date:
  - ✓ If, the difference between the dates is less than one month, the unrealized activity will be filtered with the green color.
  - ✓ If, the difference between the dates equal one month, the unrealized activity will be filtered with the color orange.
  - ✓ If, the difference between the dates is greater than one month, the unrealized activity will be filtered with the red color.

- For notifications of framework agreements and specific conventions are based on the comparison of the end date with the system date:

  - ✓ If, the difference between the dates is less than one month, the convention will be filtered with the color orange.
  - ✓ If, the difference between the dates is more than one month, the convention will be filtered with the red color.

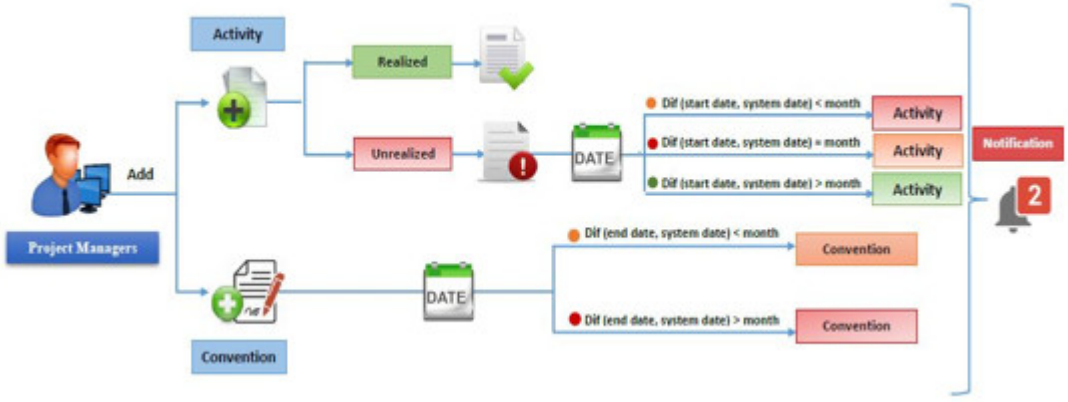


Figure 8 : warning system Process

8 Conclusion

As pressures on universities increase and the issues faced by our community become more complex, the idea of a cooperation information system can hold much promise. Through partnerships the university can contribute its part and also reap the benefits of others' high education institutions efforts. Thanks to a cooperation information system, we can accelerate learning and distribute skills and knowledge. Also, we can add depth and breadth to our scientific community impact.

Deciding on and developing a successful a cooperation information system involves a great deal of thought, planning, and coordinated effort. The SIMAcoop cooperation information system was developed after to consult with all relevant stakeholders to establish the needs for all the actors.

SIMAcoop is an application that permits to identify the members' shared vision and goals and then develop documents that outlines the terms of the partnership. SIMAcoop set up procedures for maintaining and monitoring the partnership as it evolves. It permits to detect the successes actions as well as challenges.

If setting up a partnership is important, it is more important to determine the type of partnership needed and ensure the monitoring and evaluation of cooperation actions for to maintain that there's commitment from the senior management of all partner organizations.

SIMACoop was developed by Symfony framework whohave powerful routing capabilities, annotation and regular expression usage and provides a greater choice of configuration file formats - XML and YAML. This framework template engine (Twig) allows for better template structure organization as it supports nested template blocks.

REFERENCES

[1] M. BensaouInterorganizational Cooperation: The Role of Information Technology an Empirical Comparison of U.S. and Japanese Supplier Relations Information Systems Research 19978:2 , 107-124

[2] Pinto, M. B. & Pinto, J. K. (1991). Determinants of cross-functional cooperation in the project implementation process. Project Management Journal, 22(2), 13–20.

[3] Janet R. deVry, and George Watson "University of Delaware's Faculty-IT Partnership: Educational Transformationthrough Teamwork." The Technology Source, May/June 2003

[4] Janus, Heiner and Klingebiel, Stephan and Mahn, Timo C., How to Shape Development Cooperation? The Global Partnership and the Development Cooperation Forum (March 10, 2014). Briefing Paper, German Development Institute, 3/2014.

[5] SensioLabs. The Symfony Book v. 3.1; 2016. p. 4-68. Available: http://symfony.com/pdf/Symfony_book_3.1.pdf.

[6] PHP Documentation. PHP's memory_get_included_files. Available: http://php.net/manual/en/function.get-included-files.php.

[7] PHP Documentation. PHP's memory_get_usage function. Available: http://php.net/manual/en/function.memory-get-usage.php.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Geographical Information System Tool Monitoring the Environmental Impact of Tangier Industrial Zones

Boutaina Sebbah, Miriam Wahbi, Mustapha Maâtouk
Department of Geology, Faculty of Science and Technology, AbdelmalekEssaâdi University.
Geomatics and Land Development research group
Tangier, Morocco
boutainasebbah@gmail.com

ABSTRACT

Tangier city is classified as Moroccan second economic center after Casablanca, with three classical industrial zones and two free zones. This industrial activity is expected to increase in the next few years, especially with the implementation of new industrial areas such as "Tangier Automotive City".

Although this intense manufacturing activity is essential for the city development, it has also a direct impact on the environment; one of the major changes caused by the industrial activity is the rising of surface temperatures in industrial zones compared to their surrounding areas.

This study retrieves Land Surface Temperature (LST), from three different Landsat sensors (i.e. TM, ETM+ and OLI-TIRS), using remote sensing algorithms (e.g. Land Surface Temperature, etc.) and a semi-Automatic Classification algorithm, in order to detect temperature variation in Tangier industrial areas during summer months over period between 2000 and 2016.

Results show, that the industrial zones in Tangier city (TFZ and Industrial zone of Meghogha) appear 5°C warmer than their surroundings areas, with the presence of major plants that have the highest temperature values comparing to other plants such as (YAZAKI Morocco, S.E.B.N and Delphi Packard Electric) in TFZ and (Jacob Delafon) in Meghogha industrial zone.

Keywords: Landsat, semi-Automatic Classification, Land Surface Temperature.

1 Introduction

Urbanization process is characterized by the replacement of natural surfaces by anthropogenic ones (e.g. buildings, industries…etc.) [1], [2], which alters radically the aerodynamic, radiative, thermal and moisture properties of urban areas [3],this process is related directly to the economic and industrialization development, and also to the population growth.

According to the general census conducted in Morocco in 2014, The urbanization rate in Tangier city has reached its maximum compared to other major cities (i.e. 94.34% against 90.5% in 2004)[4]

The city of Tangier (north of morocco) has also know an important modification in terms of industrial activity during the last decades, and it has become the second Moroccan economic center after

DOI: 10.14738/tmlai.54.3423
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3423

Casablanca city, with diversified industrial activity (e.g. textile, chemical, mechanical, and metallurgical industries)

Tangier has currently five industrial zones, three normal industrial zones (i.e. Industrial zone of Moghogha, Industrial zone of Gueznaya and Al majd Industrial zone) and two specific ones (i.e. Tangier Free Zone and Renault Tanger Med)[5]. In march 2016 Morocco has lunched the colossal project "Cité Mohammed VI Tanger Tech", It is a large industrial and commercial park that will involves the installation of 200 Chinese companies operating in emerging sectors such as automotive and aeronautical industry, aviation spare parts, new information and communication technologies, textiles, machinery manufacturing and other industries.

This industrial growth is indispensable for the development of the city and the entire northern region, view its positive economic and social results, but it is also accompanied by some drastic environmental degradation (e.g. air and water pollution, energy consumption… etc.).

In contrast to urban heat islands which refers to the phenomenon that urban areas tend to have higher atmospheric or surface temperatures compared to simultaneous temperatures of the non-urbanized surrounding areas [2], [6]–[8],industrial zones tend to appear as isolated “hot spots” called micro-urban heat islands [9], this elevation of industrial Land surface temperature (LST) is due in most of cases to construction materials used in industrial zones, high energy consumption and activities type.

Thermal remote sensing data have been largely used to study LST variation around the world, the combination of Geographical Information Systems (GIS) and remote sensing data can be used to provide a multi-temporal understanding of temperature change.

In the present study LST was retrieved, from three different Landsat sensors (i.e. TM, ETM+ and OLI-TIRS), using remote sensing algorithms (e.g. Land Surface Temperature, etc.) and a semi-Automatic Classification algorithm, in order to monitor and assess the spatio-temporal changes of LST in two Tangier industrial zones (i.e. TFZ and Mghogha) during summer months over period between 2000 and 2016.

2 Study Area and Data Base

Tangieris located in a relatively mounted area, the city center is located at 35°46’1.2’’N and 5°48’0’’W,it is the second largest city in Morocco with a population of about 1,065,601 inhabitants.

Tangier city has five industrial zones, twoof them were chosen in the present study in order to monitor and asses the difference of LST between thesezones and their surrounding areas (i.e.Tangier free zone, industrial zone of Meghogha) (Fig.1).

In this study, remotely sensed datasets (Level 1T products) over Tangier city were used, the data is acquired by three different thermal sensors (i.e. Landsat 5 TM, Landsat 7 ETM+ and Landsat 8 TIRS/OLI), from July to September between 2000 and 2016 Table 1,all images bands have a spectral resolution of 30 m while the thermal band have a spatial resolution of 120m, 100m, and 60m for TM5, ETM+ and TIRS images respectively.

Images with high percentage of cloudy pixels were excluded from the dataset. As a result, 22 images during the summer period from 2000 to 2016 were available (Table.1).

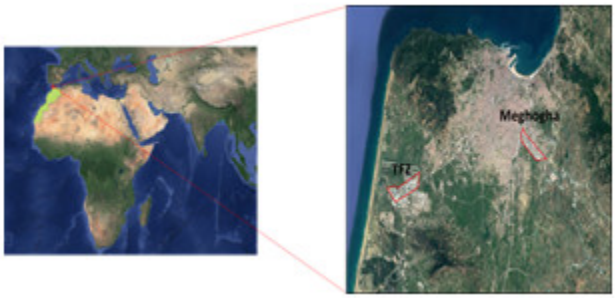


Figure 1. Study area

Table 1. Data Base Available

Satellite	*Year*	July	August	September
L7 ETM+	2000	✓	✓	X
	2001	X	✓	✓
	2002	✓	✓	✓
L5 TM	2009	✓	✓	✓
	2010	✓	✓	X
	2011	✓	X	✓
L8 OLI/	2013	✓	✓	✓
	2014	✓	X	✓
	2016	✓	✓	✓

3 Methodology

In the present paper LST was retrieved using a semi-automatic classification algorithm under an open GIS software, the hole methodology followed in this study is shown in the flow chart (Figure.2).

3.1 Conversion from DN to reflectance and at-Satellite temperature

The Digital Numbers (DN) of each band were converted to Top of Atmosphere (TOA) reflectance (1) (2) [56], then a simple atmospheric correction (i.e. Dark Object Subtraction "DOS1") was applied.

$$L_\lambda = G_{rescale} \times Q_{cal} + B_{rescale} \quad (1)$$

Where:

Q_{cal}: Quantized calibrated pixel value (DN)

$G_{rescale}$: Band-specific rescaling gain factor [(W/(m2 srμm))/DN]

$B_{rescale}$: Band-specific rescaling bias factor [W/(m2 srμm)]

$$L_\lambda = M_L Q_{cal} + A_L \quad (2)$$

Where:

L_λ: TOA spectral radiance (Watts/(m2 * srad * μm))

M_L:Band-specific multiplicative rescaling factor from metadata

Q_{cal}: Quantized and calibrated standard product pixel values (DN)

A_L: Band-specific additive rescaling factor from metadata.

For the thermal band, spectral radiance was converted to At-Satellite Brightness Temperature (3)[10], [11]. Bands 6 and 6-2 were respectively used for Landsat 5 and Landsat 7, while the TIRS bands of Landsat

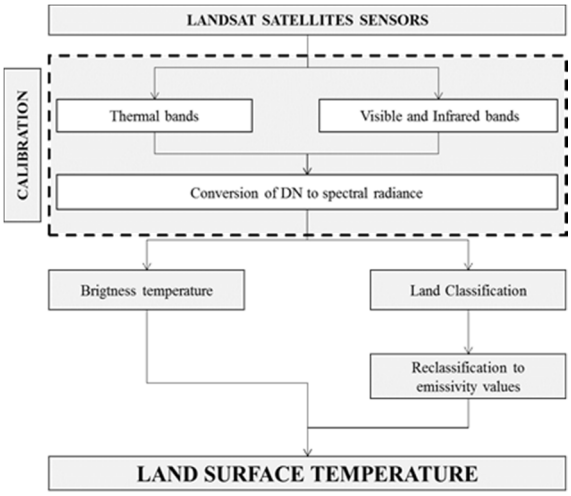
8 were designed to allow the use of split-window surface temperature retrieval algorithms, but it is recommended for users to avoid data from band 11 for the reason of the large calibration uncertainty of this band[12], thus only band 10 was utilized as a single spectral band.

After the conversion of spectral bands from DN to At-satellite spectral radiance, the thermal band of each satellite was converted to At-Satellite Brightness Temperature (3)[10], [11].

$$T = \frac{K2}{\ln\left(\frac{K1}{L_\lambda}+1\right)} \quad (3)$$

WhereK2 and K1 are two pre-launch calibration constants

K1: Band-specific thermal conversion constant (in watts/m^2 .sr .µm)



K2: Band-specific thermal conversion constant (in kelvin)
Figure 2 Flow chart of the methodology followed

3.2 Land cover classification of study area

In order to convert temperature values from At-Satellite Brightness Temperature to Land Surface Temperature, land emissivity values are required, in doing so, the study area was classified into four main classes (i.e. Built-up and minerals, bare soil, vegetation and water) basing on a supervised classification algorithm (i.e. Spectral Angle Mapping algorithm) After the classification ofland cover according four different classes, the obtained rasters were reclassified to emissivity values.

3.3 Conversion from At Satellite Temperature to Land Surface Temperature.

The Brightness Temperature values obtained from the application of (3)are temperatures that a blackbody would obtain in order to produce the same radiance at the same wavelength (11.45 for Landsat 5 and 7; 10.8 for Landsat 8). Therefore, corrections for spectral emissivity became necessary according to the nature of land cover. Each land cover class was assigned an emissivity value presented in (Table.II).The emissivity corrected land surface temperatures were computed as follows (4)[13].

$$T_s = \frac{BT}{\left\{1+\left[\frac{\lambda \cdot BT}{\rho} ln\varepsilon\right]\right\}} \quad (4)$$

Where:

Ts: land surface temperature in K;

BT: at-sensor brightness temperature in K

λ: wavelength of emitted radiance

ρ: hc/r = $1.438 \cdot 10^{-2}$ m K with

h: is the Planck's Constant ($6.626\ 10^{-34} Js^{-1}$)

c: is the velocity of light ($2.998\ 10^{8}$ m s^{-1}) and K is Boltzman constant ($1.38\ 10^{-23}$ J K^{-1})

ε: the spectral surface emissivity.

The values of λ for the thermal bands of Landsat satellites are listed in (Table. III)

Table 2. Land Surface Emissivity Values Used

Land surface	*Emissivity*
Built-up	0.94
Bare soil	0.95
Vegetation	0.98
Water	0.99

Table 3. Center Wavelength of Emitted Radiance of Each Satellite

Satellite	*Band*	*Center wavelength*
Landsat 5 and 7	6	11.45
Landsat 8	10	10.8

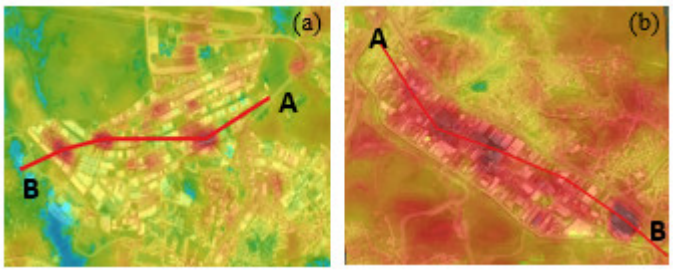


Figure3 Travers pointes of the two studied area; (a) TFZ , (b) Meghogha

3.4 Calculation of LST variation in industrial zones

In order to study the difference of LST between thetwo industrial zones (i.e. TFZ and Meghogha) and their surroundings areas, travers points over the studied areas were established (Fig. 3). The values obtained from each traversed pixel were used to plot the variation curves in each studied area from 2000 to 2016, as well as the average curve of LST variation during these 16 years.

4 Results and discussion

LST maps (Fig.4, Fig.5) show that temperature values in the studied areas vary between 28 and 42°C, temperatures values differ from season to another, with the presence of cold seasons such as summer 2002,2011 and 2014 where LST values vary between 28 to 36°C and also some warm seasons characterized by temperatures values that run between 36 and 42°C like summer 2000, 2000 and 2016.

In the first studied area (i.e. TFZ) and before the implantation of the industrial zone, the area tends to appears homogeneous in terms of LST values, from 2002 onwards it was noted the appearance of three major plants (i.e.YZAKI, S.E.B.N and Delphi) that appear as hot spots; they have in general temperature higher than thesurroundings area.

In the graph of LST variation (Fig.6); temperature varies in this areas between 28 and 43°C, the three pics in the graph present LST values of major plants already cited. LST values have attend their maximum during 2013 which was a

relatively hot season; while the minimum values were recorded during summer of 2014, it is remarkable that the temperature between these pics decrease, and also when moving away from the TZF to the surrounding areas the LST decrease to 31 °C, note that these three factories that participate in the heating of TFZ area and also its environs are wiring plants specialized in the production of cables.

InMeghogha industrial zone, which is the oldest industrial area in Tangier city, results show that LST in this area is relatively higher than its environs, temperature values vary between 31 and 45°C; with the presence of a major hot spot present 'Jacob Delafon', a plant which is specialized in the fabrication of sanitary equipment. LST graph in Meghogha industrial zone (Fig.7), results show that the maximum temperature values were recorded in summer 2013 while minimum values were recorded during summer 2014.

In both studied areas results shows that the major plants appear as hot spots even during the coldest seasons or the warmest ones, they appear 5°C warmer than their surroundings areas.

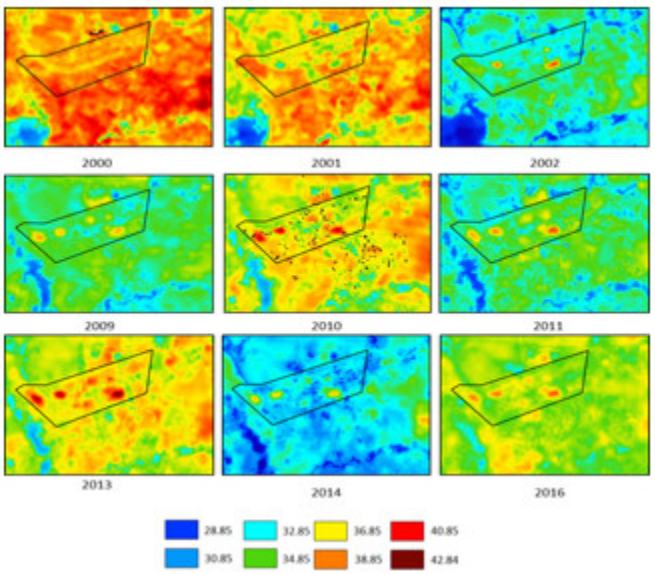


Figure 4 LST maps of Tangier free zone

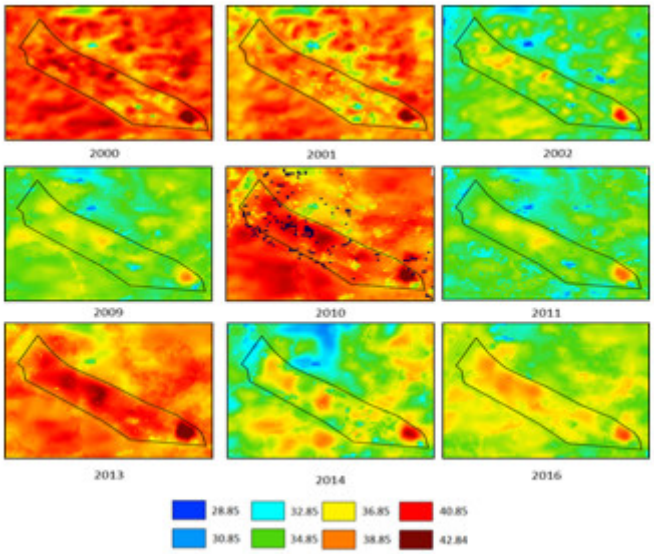


Figure 5. LST maps of Meghogha industrial zone

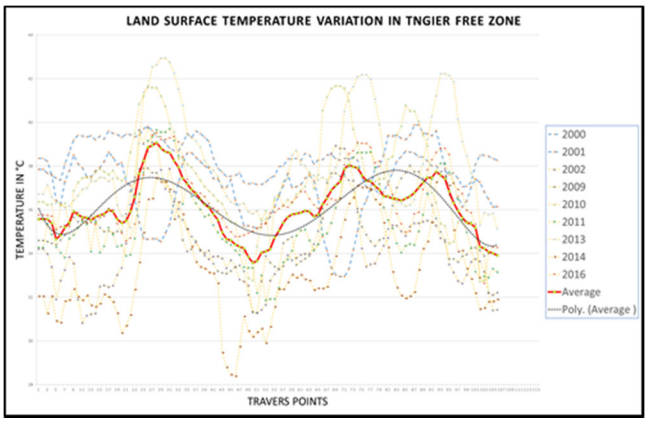


Figure 6. LST variation in Tangier free zone from 2000 to 2016

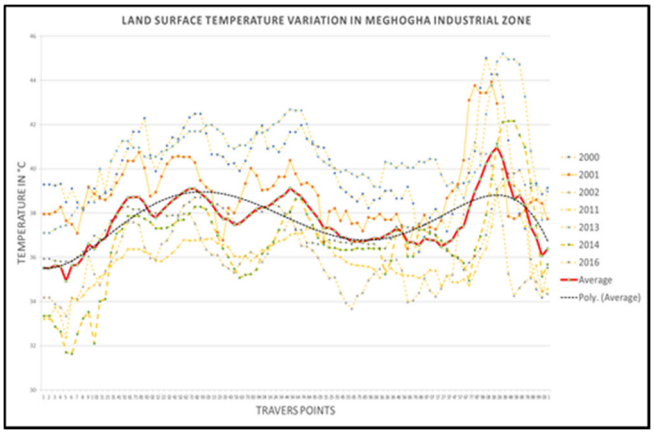


Figure 7. LST variation in Meghogha industrial zone from 2000 to 2016

5 Conclusion

This study has used a combination of remote sensing data and GIS tool to evaluate LST variation in two industrial zones of Tangier city, the obtained results has shown a difference of 5°C between industrial

zones and their surrounding areas, with the presence of some micro urban heat island presenting the major plants in industrial zones.Given the climate changes that induce temperatures raising, this phenomenon of micro urban heat islands could be more widespread in the surrounding areas and may have some drastic consequences, in particular citizens health status.

Thus,this industrial activity could not be stopped, view its primordial role in city development, but it is necessary to take into account the environmental aspect, in order to protect the surrounding areas, and reduce the temperature in industrial zones allto achieve a sustainable development of the city.

REFERENCES

[1] R. Anniballe, S. Bonafoni, and M. Pichierri, “Spatial and temporal trends of the surface and air heat island over Milan using MODIS data,” Remote Sens. Environ., vol. 150, pp. 163–171, Jul. 2014.

[2] J. A. Sobrino, R. Oltra-Carrió, G. Sòria, R. Bianchi, and M. Paganini, “Impact of spatial resolution and satellite overpass time on evaluation of the surface urban heat island effects,” Remote Sens. Environ., vol. 117, pp. 50–56, Feb. 2012.

[3] J. A. Voogt and T. R. Oke, “Thermal remote sensing of urban climates,” Remote Sens. Environ., vol. 86, no. 3, pp. 370–384, Aug. 2003.

[4] Haut Commissariat au plan, “Population legale des regions, provinces, prefectures, municipalites, arrondissements et communes du royaume.” 2014.

[5] “Royaume du Maroc - Centre d’Investissement de la Région Tanger Tetouan Al Hoceima.” [Online]. Available: http://www.investangier.com/.

[6] M. L. Imhoff, P. Zhang, R. E. Wolfe, and L. Bounoua, “Remote sensing of the urban heat island effect across biomes in the continental USA,” Remote Sens. Environ., vol. 114, no. 3, pp. 504–513, Mar. 2010.

[7] D. Zhou, S. Zhao, S. Liu, L. Zhang, and C. Zhu, “Surface urban heat island in China’s 32 major cities: Spatial patterns and drivers,” Remote Sens. Environ., vol. 152, pp. 51–61, Sep. 2014.

[8] J. Quan, Y. Chen, W. Zhan, J. Wang, J. Voogt, and M. Wang, “Multi-temporal trajectory of the urban heat island centroid in Beijing, China based on a Gaussian volume model,” Remote Sens. Environ., vol. 149, pp. 33–46, Jun. 2014.

[9] C. Aniello, K. Morgan, A. Busbey, and L. Newland, “Mapping micro-urban heat islands using LANDSAT TM and a GIS,” Comput. Geosci., vol. 21, no. 8, pp. 965–969, Oct. 1995.

[10] G. Chander, B. L. Markham, and D. L. Helder, “Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors,” Remote Sens. Environ., vol. 113, no. 5, pp. 893–903, May 2009.

[11] United States Geological Survey, “Landsat Missions: Using the USGS Landsat 8 Product.” 2015.

[12] United States Geological Survey, “LANDSAT 8 (L8) Data users handbook.” 2016.

[13] Q. Weng, D. Lu, and J. Schubring, “Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies,” Remote Sens. Environ., vol. 89, no. 4, pp. 467–483, Feb. 2004.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

The Rate of Local Integration in the Supply Chain of A Car Manufacturer

Raissouni Rajaa, Hamiche Mhamed
Laboratoire de Gestion, Economie et développement durable
Université Abdelmalek Essaadi
Tanger, Morocco
rajaaraissouni@gmail.com, Hamiche2020@gmail.com

ABSTRACT

The rate of local integration is the share of production produced in Morocco in a context of production broken out around the world by multinationals When all the parts of an automobile are imported into Morocco and then assembled on the spot before being re-exported in the form of whole cars by a builder, then local integration comes down to assembly

Local integration is a permanent challenge for Morocco's industrial policy. It is about bringing into the realm not a builder or two, but all their chain of subcontracting. "The challenge is to become the backbone of Europe for sourcing", and according to Renault Maroc, this sourcing "integrates local purchases and also export (excluding vehicles)". It is understood that the project aims to position Morocco as an exporter of automotive components of first rank. Moreover, the group is a VRP in Morocco since it has hosted 250 international suppliers in the Tangiers plant "to present the new ecosystem, the industrial platform of Tangiers, to encourage them to invest in our country." It is also learned that the local integration platform that will be created will include purchasing, engineering and logistics resources in order to optimize local sourcing and to make Morocco a solid industrial export base.

In this article, we question the impact of local integration on the supply chain of Renault Morocco, which leads us to ask the following question : Does the increase in the local integration rate reduce the purchase costs of the multinational Renault?

In this regard, we adopted a two-step approach: First we focused on the conceptual aspects of the purchasing function and the theoretical basis of sourcing. Secondly then presented a project for local integration and diagnosis of the body & Electrical perimeter applied within the company Renault Morocco.

Keywords—Supply chain; Local integration rate ; Automobile industry; Sourcing procedures; sourcing practice.

Important note :The amounts and prices reported in the article are approximate figures and are not actual prices.

DOI: 10.14738/tmlai.54.3425
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3425

1 Introduction

The automobile industry has become Morocco's leading export market and one of the most successful industrial pillars in the Moroccan industry. First with the establishment of Renault Tanger in 2008, which is the biggest foreign investment in Moroccan history, furthermore with the anticipated arrival of Peugeot-Citroën (PSA) in Kenitra in 2019; therefore, Morocco remains an attractive country for major car makers due to its political stability and favorable geographical position compared to its neighbors in the region.

The development of a solid industrial fabric is the result of further integration of the automotive value chain in Morocco. Indeed, the year 2016 testifies to the signing of supplier agreements called Ecosystem, mobilizing investments of the order of 10 billion dirhams.

These agreements were signed in the presence of His Majesty King Mohamed IV and several ministers as well as the Director of Operations, of the Renault group Middle East and India branch. This project involves developing a global supply platform that generates a figure of 20 billion dirham per year. It will contribute to the creation of 50,000 new permanent jobs with the objective of doubling this number by 2020 to more than 160000. In short, the ecosystem signed in April 2016 includes three types of agreements:

- A framework agreement on the establishment of the Renault ecosystem.
- An industrial property agreement.
- A convention on executive training and vocational training in the automotive sector.

Morocco commits to give subsidies to investments and tax advantages to suppliers who intend to settle within the country. Renault took advantage of this ecosystem agreement by increasing its local integration rate and consequently reducing and optimizing its logistics and purchasing costs.

Generally, the purchasing costs of multinationals such as Renault represent a significant percentage of the cost of their products. To this end, they still hope to optimize them by several levers, local sourcing is a strategic tool to achieve this objective through its direct contribution to the purchasing performance and consequently the company's performance.

2 Literature review: Conceptual aspects of the purchasing function and the theoretical basis of sourcing

2.1 Conceptual Aspects of Purchasing

2.1.1 Definition of purchases

The concept of purchase has no universally known definition. This is the reason why there are several variants of the purchase functions in companies. Besides, it would be easy to generalize this definition by defining the purchasing function as "Function responsible for the acquisition of the goods or services necessary for the proper functioning of the enterprise". More precisely, the purchase's mission is to acquire the goods and services required, at the desired quality level, in the desired quantities, within the expected timeframes and in the best conditions of service and security of supply .

The purchasing function contributes directly to the company's corporate strategy. Today, it is difficult to increase profit by sales, but it is possible to improve a company's productivity by reducing costs and improving the technical definition of the product.

Organizations have long neglected the purchasing function. Yet purchases account for an average of 70% of the turnover of industrial companies. Globalization has reinforced the impact of the purchasing function on the company's performance in terms of quality, cost and time.

Indeed, the purchase function is the intermediary between the internal services and the suppliers. In order to meet the specific expectations of each department of the company (production, logistics, development, etc.), the buyer is obliged to know the characteristics of suppliers and their products and services requested. In order to be as efficient as possible, it is important to keep a watch on the market to know its trends and the various suppliers present and likely to satisfy the need of the company.

In addition, the purchasing function differs categorically from that of supply. This problem of confusion and mixing has always been questioned by researchers. For this, and before projecting into the fundamental aspects of the purchase, an explanation of the difference between the supply and the purchase is necessary.

2.1.2 The economic and strategic stakes of the purchasing function:

Economic issues

The preponderant role played by business purchases is due in large part to a favorable economic and social environment. The main justifications that could be mentioned are:

- A growing global economy.
- An improvement in the living conditions of individuals.
- The development and the explosion of new information and communication technologies (ICTs).
- The transition from a production economy to a market economy.

Sustained economic growth

Business activities were revived with the return to growth. This growth is reflected in several ways. First, an increase in demand, in terms of quantity but above all in terms of quality, has forced companies to increase their productivity. This situation leads companies to constantly seek the best economic, technical and industrial policies to be able to satisfy this demand.

Secondly, we have the phenomena of globalization and internationalization that bring companies, notably through mergers and acquisitions, to integrate new markets and thus find new outlets for their products. However, it should be stressed that this economic growth also leads to increased competition. This is becoming so harsh that only companies that are innovative, adaptable and competitive can survive to meet the demands and quality needs of ever-increasing and increasingly demanding consumers.

The development of NICTs

With NICTs, especially the Internet and Intranet, there is an increasingly strong structuring in the management and dissemination of information internally. Thanks to these technologies, companies have new tools allowing them to constantly update the data but also to disseminate it to all the employees concerned, which allows a collaboration almost in real time.

Transition to a market economy

To these first observations, we can add the fact that over the past two decades we have gradually shifted from a production economy to a market economy. But we will return to this change in more detail because

it is one of the factors explaining the evolution of the purchasing function. Faced with all these factors, companies have had to create new capabilities that allow them to generate productivity gains and significantly reduce their costs to compete while satisfying their customers. It was on this principle that the procurement function was created.

Strategic issues

The Search for Better Competitiveness

International competition is forcing all companies to increase their competitiveness. To do so, during the 1970s, the traditional strategic lever was to develop commercial investments, in conjunction with the various productivity actions implemented in factories. These measures have often been effective, but the gradual stabilization of results has led the Directorates-General to move towards an area that is still under-exploited: Purchasing. The main reasons are:

- The share that the purchase items represent in the income statement. If we consider that beyond the raw materials and components, the various forms of subcontracting, industrial services of all types are purchases.
- For the last 10 years, large firms have often chosen to focus on one (or some) core business and thus abandon industrial activities
- Previously integrated into certain sectors, in addition, the purchase of complete functions is gradually replacing outsourcing: an increasing share of value added is transferred to suppliers
- Finally, purchases constitute an area in which gains (savings) are made in exchange for low investments (mainly men and information system): the leverage effect is interesting and savings are quickly Income statement.

Developing the Mastery of Technologies

The second key factor of competitive advantage is the development and management of a technological heritage. Almost always, the company does not have sufficient human and material resources to master all the technologies necessary for the development of its products. Moreover, in the high-tech sectors, the high cost of study and development equipment, as well as that of industrial processes, make it impossible to depreciate them on even average quantities and impose specialization in industrial sectors. Thus, the field of procurement of industrial policy to be carried out in relation to suppliers becomes a new strategic factor.

Mastering High Quality and Providing High Level Service

Quality assurance and quality control procedures for purchases impose a particular requirement in that they involve the implementation of various external audits and the steering of suppliers in the framework of permanent progress plans.

As for the notion of service, it is expressed mainly by the implementation of a reduction of the time, in order to minimize the dysfunctions and to gain in reactivity. Once this objective has been achieved internally in the plant, as well as in the downstream distribution system, the interfaces with the suppliers have to be developed, and this belongs to the purchasing domain.

Minimize Development Cycles

Nowadays, winning companies are the ones that manage to reduce all cycles, from initial product design to delivery to customers (time-to-market concept).

Thus, to reduce industrial delays, there are also actions to optimize the design and development processes of new products. Those who do so can afford to renew their product range faster than competitors, and can thus shorten their lifetimes.

2.1.3 The purchase cost

The purchase cost is made up of everything that the element cost, excluding recoverable taxes, from its conception until its entry in the store or its put in use. It is therefore composed of:

- The amount shown on the purchase invoices net of trade reductions and recoverable taxes;
- Of all costs related to the purchase (ancillary costs to purchase) such as third party transport, customs duties, commissions and brokerage and installation costs ...;
- Possibly, procurement costs, some components of which have problems in their allocation (acquisition, receipt, handling, accounting, etc.).

The determination of purchase costs is essential for the good, nature of goods or materials or supplies. It is for the purpose of enabling the recording in the accounts of fixed assets or stocks in accordance with the directives of the accounting law of 1983.

2.2 Sourcing strategies

Sourcing is the research, location, selection and evaluation of a supplier in order to meet an identified need. However, in a context of globalization, multinationals manage sourcing around the world by managing a global supply chain for their production sites scattered across the globe. To do this, multinationals seek to reduce the logistical costs of their factories by looking for locally installed suppliers who can supply and deliver according to their specific standards.

2.2.1 Sourcing strategies according to the supplier panel

The different sourcing strategies

Sole sourcing

It is a single source of supply for one type of product. It can be the result of sourcing from a supplier whose production capacities are rare in the market; But it would be wise to look at the supplier market to compare the supply of other suppliers with the one already contracted.

The use of a single supplier presents a strategic risk insofar as an incident can lead to a break in the supply chain, the consequence of which can be damaging for the enterprise. The link between sole sourcing and supplier relations gives rise to a less formal and longer-term contact. But having a sole sourcing strategy to increase quality is not always true.

Dual sourcing:

It involves using two suppliers to produce the same component or buy the same raw material. In general the business is shared for 70% with the supplier Y and 30% with the supplier X. It is in some way a response to the risk of the single source of supply since the supplier acts as a backup in case the other would be

failing. Many companies consider dual sourcing as important because it allows the company to have some flexibility as well as avoid being frozen by long-term contracts with a single supplier.

Multi sourcing:

Multi sourcing is when three or more suppliers are used by the company for the same product. This corresponds to a strategy of diversifying sources of supply when, for example, a parent company orders from a number of suppliers close to its factories and markets a similar product to supply the local demand.

Back-up strategies

Cross sourcing:

Cross sourcing is a sourcing strategy in which a company uses a supplier in the production of a good or service for a production domain of the company and another supplier producing the same good or service for another area of production line.

The idea is to have a back up supplier and develop a competition between the two suppliers for future contracts to be awarded. Competition between suppliers is a means of encouraging emulation but should not be an instrument of pressure of the client company on its supplier.

Suppliers of fluctuating demand

Some industries experience a peak demand during a season of the year. They use a supplier that produces continuously during the year. During the strong period, one or more other suppliers will support the peak demand. The degree of strategic relationship with the temporary supplier is not necessarily very high.

Sometimes it serves as a "spare wheel", thing that hampers to carry out a strategic purchasing policy with a reduction in prices and the appearance of innovation.

2.2.2 The geo-location of the company and the location of resources

Ocal sourcing

Local sourcing involves working with regional or national suppliers, defined according to the location of a business activity as a factory. Local sourcing has the advantage of reducing the risks associated with logistics. Another advantage is to benefit from a product delivered with speed, when the freshness of the products is critical or when the demand evolves quickly.

For a company that does not internationalize its sourcing, local sourcing provides, at a lower cost, a developed network of partner suppliers to cover the demand in either local or international markets. However, if these sourcing concerns manufactured product in a declining cycle, low-end, designed by unskilled employees, the local sourcing in certain regions is not advantageous because costs become the first criterion.

International sourcing

International sourcing is possible through the development of transport and telecommunications which makes it more accessible. This international sourcing is used to benefit from a comparative advantage like that of Silicon Valley where innovations are strong because emulation is favored. Thus appears a geography of sourcing where the firms look for quality in some leading countries like France, Japan, Germany and the United Kingdom .

International sourcing has developed with the emergence of competitive players outside the borders. With the collapse of the Soviet Union, China's openness, WTO agreements, the development of NAFTA, raw materials, plastics, electronics are bought more frequently from international sources.

This sourcing is adapted when, for manufactured products, the time taken for ship delivery is not paramount. Otherwise the airway remains a possibility to transport in the emergency but its use is very expensive. International sourcing requires taking into account the total costs of delivery (transport, customs clearance, banking, insurance, quality inspection); Trade regulations, both restrictions and incentives; Time to Market (flexibility); Value-added services (a service which is offered by a supplier and which makes it possible to dispense with one more intermediary); Internet communication tools. The barriers are understanding of international procurement procedures (import licenses, certificates of origin), extended delays, cultural differences, currency risks. Finally, having international suppliers has the effect of increasing the inventory level resulting from a policy of risk reduction (downtime, terrorism, change, policy) that leads to the accumulation of security stocks.

2.2.3 The sourcing procedure

Expenditure Analysis, Spend analysis

Strategic sourcing begins with spend analysis up to the procurement automation process . Spend analysis refers to cost management. It is a way of analyzing cost structures that can be carried out by product or by product groups. The products purchased and the procedures leading to these purchases are analyzed in great detail.

The proactive specifications

Before a sourcing project, the company establishes a general project plan (mapping) in which it formulates its needs, proposes modes of action in the supplier market, makes internal recommendations, identifies the cost sources (cost drivers). It is the stage of constitution of the proactive specifications. This is the preparation stage prior to the decision to attack (as in marketing) the dense supplier market.

Even before the detailed specifications have been drawn up, it must precede a strategic decision by the company's top management. The intention is specified, communicated and comes into application by the constitution of a sourcing team. The latter will attempt to determine the buying segments where the proactive approach to the supplier market can be carried out. Once defined, the team draws up the complete specifications by consulting the other internal clients in order to guarantee the general approval in a strategy of global strategy known to all.

Data collection and processing

Any sourcing project, whatever its strategic and geographic orientation, requires the company to collect and analyze all kinds of data in order to prepare the sourcing process and anticipate the risks. This may be data from previous sourcing operations: Results, quality of products purchased, and respect of deadlines by the supplier. This may include data on the economic environment such as inflation rates, world commodity prices, exchange rates and the transport network. Also the buyer is required to learn about the legislative environment. The latter can be very changeable and contain legal gaps in some countries; It is necessary for the firm to guard against possible risks due to non-protection by the law, such as intellectual property.

Generally it is necessary to collect the strategic information that will help to establish the sourcing decision, indeed it is necessary to look for 3 important information:

- Geo-location of supplier: this information will allow defining the logistic flows and the daily rates of delivery.
- Quality of supplier and price: An ISO certification of supplier is very important information that proves the quality of the products.
- Production capacity: knowing the supplier's production capacity will make it possible to know its competitiveness and its flexibility towards an unpredictable increase in demand, which generates large investments.

The offers received and selection suppliers

Supplier selection is a critical phase in the sourcing process, as it has a significant impact on the company's performance.

Indeed, it is necessary to develop a network of reliable and competitive suppliers so that the company can raise and survive in a competitive environment. However, in order for the company to be able to provide a quality product at a reasonable cost, it must virtually monitor the performance of its suppliers.

However, there are two major problems:

- The determination of the number of suppliers and the mode of relationship with them: this choice is largely influenced by the strategic activity area of the company as well as its management mode,
- The selection of the best suppliers: this problem arises after the determination of the number of suppliers selected, however we will explain in detail in the next section the different techniques designed by researchers in the determination of the best suppliers and the fundamental criteria in the decision of making supplier appointment .

3 Sourcing practice, analysis and development of local integration

3.1 Host organization

Renault Tanger extends over 300 hectares as an assembly plant with a production capacity of 400,000 vehicles per year. 90% of the vehicles produced will be destined for export. This site complements Renault's industrial system for low-cost vehicles derived from the Logan platform. It is operational since the beginning of 2012. This plant with a capacity of 400,000 vehicles per year starts with the production of two new entry models: the Lodgy family and a small commercial vehicle named dokker, also declined in a particular vehicle version. This site will increase the volumes and expand the range of the M0 range.

Thus, the Renault Group will be able to meet the sustained demand from customers for entry-level vehicles, renowned for their unparalleled performance / price ratio.

In addition, the site is based on the strategic position of the port of Tangier between the Atlantic and the Mediterranean, a developed and competitive fabric of suppliers, and a workforce trained in the best automotive techniques.

The Renault Nissan factory in Tangier was inaugurated on February 9, 2012 by his majesty King Mohamed VI and Carlos GHOSN, President and CEO of the Renault-Nissan Alliance. L'Usine has direct access to the

port platform of Tangier Med port. It is also the first automotive factory in the world that has zero jet industrial fluids.

Table 1 : Fact sheet of Renault

Fact sheet

Social reason	Renault Tanger Exploitation SA
Legal form	S.A.S
Director	Paul carvalho
Number of employees 2016	332 Frames, 890 Employees, 5900 workers, 1065 interim
Address	Renault Tanger exploitation, Free zone of Melloussa, Tangier
Manufactured products	Lodgy J92 Dokker FK67 Sandero X52
Factory Certification	Factory 100% zero emission
Area	300 Hectares, including 220 hectares of covered building
Creation date	16 January 2008

Table 2 : Key figure for Renault

Key figures

Number of assembly lines	2 lines of assembly: Tangier 1 and Tangier 2
Area	300 Hectares, including 220 hectares of covered building
Creation date	16 January 2008
Production	400,000 vehicle. / Year at term
Number of employees	7000 at the end of January 2016
Factory Certification	Factory 100% zero emission
1st phase of the project	30 vehicle / hour, 200,000 vehicle / year
2nd phase of the project	60 vehicles / hour, 400,000 vehicles / year
Legal form and distribution of capital	S.A. Caisse de dépôt et de gestion 47.6%, Renault SAS 52.4% du capital de Renault Tanger Méditerranée

3.2 Presentation of the local integration project and diagnosis of the perimeter body & Electrical

3.2.1 Presentation of the local integration project and its role in improving purchasing performance

General presentation: Local integration of automotive parts

This trend is motivated by the gains to be made in terms of geographic proximity, lower logistics costs and exchange rate risks. To benefit from the cost advantage, Renault Maroc is looking for local partners for business opportunities in order to establish a Win-Win relationship and also minimize logistical costs.

The supplier's competitiveness remains the main concern of Renault Maroc. To this end, it accompanies these suppliers with the ANPQP (Alliance New Product Quality, Procedure) approach, which is founded by Renault-Nissan to make annual agreements and provide specialized teams for suppliers to transfer know-how, especially in terms of quality management and competitive production.

The ANQP is the common procedure to cover all the quality assurance activities expected for the POE (Outsourced Product of the Renault plants). It was developed to define the common requirements of Renault and Nissan towards their suppliers from the initialization of the project, through the manufacturing agreement, to the end of production series. The challenge is therefore to reach the objectives qualities, cost, delivery time and overall expectations of the customer.

The problem that generally arises for the design of automobile parts in Morocco is the lack of technology and experience. Indeed, the Moroccan industrial fabric does not have the technologies necessary for the manufacture high quality parts or the human skills necessary for the realization of a project of high posture. However, there are multinationals, established in Morocco to fill the gap, and already part of the Renault-Nissan supplier panel.

The criteria requested by Renault-Nissan Purchasing Organization (RNPO) is a joint purchasing organization between Renault and Nissan, which was set up in 2001 and has gradually evolved to cover 100% of purchases since 2009. Common purchases Mean that Renault and Nissan have mandated RNPO to build a business panel and monitor the performance of suppliers), for the choice of suppliers is grouped in 4 primary parts:

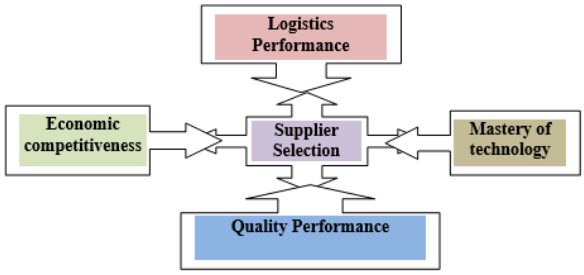


Figure 3: Supplier Choice Criteria at RNPO

Local integration is an effective tool for cost reduction, but the process is slow and requires heavy investment of tools and studies. A profitability study is carried out at the end of the consultations to decide on the profitability of the project for Renault.

Today, the rate of local integration is 33%, which remains very low compared to the targets set at the start of the Renault Tangier plant.

Indeed, the last supplier agreement signed in April 2016, and the establishment of ecosystems, will allow Renault to boost this rate of local integration thanks to permanent support from the Moroccan State,

specifically the Ministry of Industry for The development of the industrial fabric in Morocco. The objective is to achieve a local integration rate of 65% by 2023, which remains a challenge for both stakeholders namely Renault and the Moroccan State. So the local integration project takes into account not only the optimization of costs but also the commitment vis-à-vis the government.

SWOT analysis of local integration

The SWOT analysis of local integration will enable us to identify, on the one hand, the components on which we can rely and count, and on the other hand, the weaknesses that need to be reinforced. Here is the SWOT analysis:

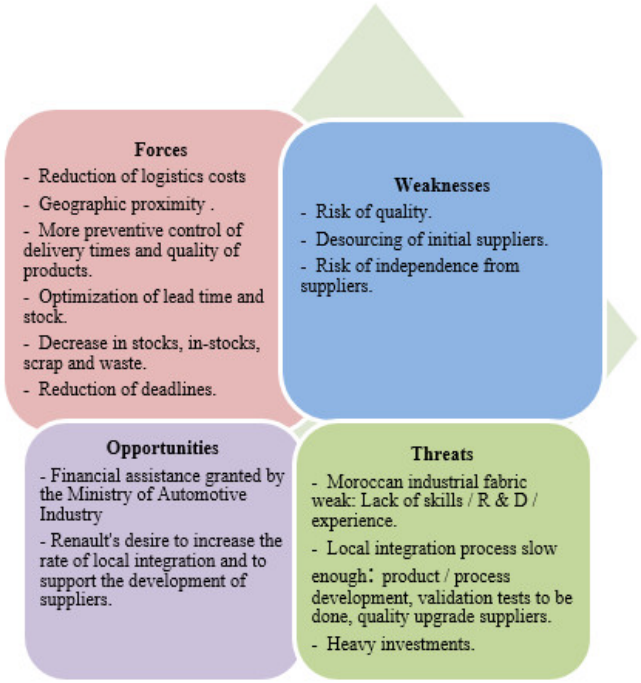


Figure 4: SWOT Analysis of Local Integration

Local integration steps

The process of local integration takes place in several stages, from the determination of the perimeter impacted up to the appointment of the supplier, which makes sense, since sourcing is a flexible process that changes according to sector of activity and the nature of the business. Below the different actions carried out for each stage of local sourcing:

Step 1: Determine the scope of action:

In this phase, the following elements are specified:

- Identification of the parts concerned.
- Identification of projects.
- Identification of customer sites.
- Determination of current suppliers.

- Identification of diversity (All existing references), and calculate the mix of references by several criteria either by volume, power train, or product range, the choice of mix depends on the nature of the part.
- Location.
- Prospecting and determination of suppliers to consult (visiting local suppliers, consulting current foreign suppliers for a possible establishment in Morocco).

Step 2: Selection Criteria:

In this phase, we determine our supplier selection criteria, which are: Geo-location (Integrated multinational suppliers)

- Economic competitiveness.
- Technological mastery.
- Logistics performance.
- Quality performance.

Step 3: Define local integration scenarios:

After defining the selection criteria, it is recommended to draw a road map for possible scenarios of the local integration.

Step 4: Recovery of engineering definition:

Engineering prepares the technical definition in parallel with the purchase for 3D modeling and 2D of the part.

Step 5: The sending of consultation files to the supplier (Request for information):

The purchase sends the RFI file. This is the equivalent of the purchase specification.

Step 6: Kick-off meeting with suppliers:

The purpose of this meeting is to discuss the investments necessary for the start-up of the manufacture of the part.

Step 7: Launching of logistics costing :

To calculate the necessary transport costs, and compare the new and the old logistic path.

Step 8: Comparative Analysis, Supplier Returns:

A comparison of the offer of new supplier and the offer of the current supplier, it is done on the following points:

- Price comparison.
- Comparison of logistics costs.
- Comparison TEI (Investment Ticket).
- Comparison of delivery times.
- Comparison of quality.

Step 9: Purchase Negotiation:

This stage of negotiation is essential for building a lasting relationship with the supplier.

Step 10: Appointment of Supplier:

The supplier is appointed for the delivery of the part.

Step 10: Appointment of Supplier:

The supplier starts the manufacturing process of the part.

Finally, this step is the standard procedure for a local sourcing project, the order of these steps differs from one project to another.

Before starting the project of our integration we will first analyze the inventory of the perimeter purchase Body & Electrical.

3.2.2 Analysis of the state of the premises and diagnosis of the perimeter body & Electrical

Inventory of fixtures

The starting point of the project is to do an analysis of the Body & Electircal perimeter of Renault Maroc. The perimeter that will be analyzed corresponds to everything related to the body of the car and the electronic elements and systems that constitute it.

The Body & Electircal perimeter is made up of 5 buyers, each one dealing with a number of parts and their suppliers. Some of these parts are locally integrated and another part is not. To do that, we tried to develop a supplier panel so that we could have more visibility.

In this perimeter there are 4 families of purchases, below a table that summarizes all the types of purchases in this perimeter:

Table 5: on the Body & Electrical perimeter parts panel

Body Equipment Purchasing	- Glazing. - Watertight seal. - Wing mirror. - Lighting. - Wiping.
Purchases climate and cooling system	- Air conditioner. - Radiators. - Radiator hoses. - Compressors. - GMV: Ventilation motor unit.
Purchasing Multi-Media and Mechatronics	- Clusters. - Door mechanism. - Hinge. - Switch. - Combi switch. - Door switches. - Electronic keypad. - Probes. - Speakers. - Antenna. -Microphone and displays. - Radio / Navigation system.
Buying Harness, Energy Systems	- Wiring. - Drums. - Starter.

Indeed, five buyers share among them these parts, their management, their supplier panel, and supply monitoring.

Indeed, this perimeter is very strategic, since it gathers a very interesting percentage of all parts worked on vehicles. That said, buyers must have background and training in the technical field and specifically, in mechanical and electronic engineering to handle negotiations very well with its supplier.

We have chosen the group of merchandise climate and cooling system, for a possible local integration, however before beginning the consultations, we have made a small diagnosis of this grouping of merchandise to study the parts not yet integrated locally and those potentially to be there in 2016.

Diagnosis and analysis of the parts group Climate and Cooling System

We have created a table that includes all the families of purchases and the parts allocated to them. Subsequently the document was sent to the buyer who is responsible for the perimeter climate and cooling, so that he can tell us which purchasing families he is responsible for, their suppliers solicited and the parts that are integrated locally and have not yet been.

This support will help us to have a visibility on the parts that can be integrated locally and to identify it in order to launch the supplier consultations on these parts.

Table 6: analysis of the climate and cooling system

The parts of the climate perimeter and cooling system		Buyers	Local integration 2016	Suppliers	City	Integrated Potential 2017
Climate And cooling system	Compressors		No			Yes
	HVAC manual		Yes			
	HVAC regulated		Yes			
	GMV		Yes			
	Air guide		Yes			
	Radiator		Yes			
	Radiator hoses		No			Yes
	Hoses for water evacuation		Yes			

This table is a summary on the integrated parts and the others that are potentially integrable for the year 2017 and for which we already have an idea on the feasibility of the project. This table contains the following:

- Suppliers solicited for each commodity group.
- The parts that will be integrated for the year 2016.
- Potentially integrable parts for the year 2017.

After drawing up this table, the engineering department and the project purchases made a meeting to choose between the compressors and the radiator pipes for a possible local integration. We found that there is no international supplier installed in Morocco that would have the necessary technology to produce compressors of good quality. In the end Renault chose to start on the local sourcing of radiator pipes.

3.2.3 The project framework

We concluded that SMA will be interested in the supply of radiator pipes as shown in the table above shows that Renault has already integrated locally at SMA Casablanca . The water drain pipes that are more or less similar in their production than the radiator pipes, in addition SMA already has an experience effect in the production of radiator pipes since they produce this type of parts in its plants everywhere In the world for Renault and for other car manufacturers.

The choice of radiator pipes was validated by the engineering department, as well as the purchase projects including with the approval of Chamberlain Nicolas project manager local integration Renault Morocco.

The supplier SMA with its new site in Casablanca represents an opportunity to be seized by Renault. The table below summarizes the different levers of this project:

Table 7: Project Framework Summary

Target Supplier	SMA Morocco
Actual Supplier	INAS India
Year of projection	2017/2018/2019
Turnover	400 000Euro
Projects	Lodgy, Sandero, Dacia Logan, Docker
Impacted perimeter	**Radiator Inlet Hose** **Radiator Outlet Hose** **Heat Inlet Hose** **Hose outlet radiator heating** **Hose intake tank degassing** **Hose outlet tank degassing** **Hose purge**

A precision, to clarify the confusion between the nature of the part and the references. Each part has a number of references, in the case of this project there are 800 references for the family of radiator pipe parts. The reference is a code which designates the technical characteristics of the part.

3.3 Consultation phase

Generally this phase is subdivided into two main components:

3.3.1 The development of the purchase specifications

In the Renault standard the purchase specification is constituted in the form of an RFI, Request for information. Indeed for the realization of the mission it is necessary to collect different information from different departments so that the file is in good and due form.

The volume of vehicle projects

At Renault each vehicle is associated with a specific code and a project manager responsible for the monitoring of the sourcing carried out. To do this consult the project managers purchase to have the volumes impacted by our sourcing on each type of vehicle namely: Docker, Dacia Logan, Sandero, lodgy.

Indeed, the designation of the volumes for the consultation is done according to a well-defined purchase strategy, to simplify, the project managers buy have two different volumes for each project, the first is the average volume and the second c 'Is the maximum volume, this difference is due to the unpredictability of the scenarios throughout the duration of the project.

Generally, consultations are conducted on average volumes. This is what we did, for the simple reason that if the supplier is competitive on the average volume, it will automatically be competitive on the max volume.

Mix volumes and references

The second information to be included in the RFI is the Mix volumes and references. The Mix volume is the criterion for which the supplier must return the volume of the previous table. In addition, since we do sourcing to radiator pipes, we must consider the types of motorization and the different mechanical aspects of the parts.

This information is retrieved from the engineering department. They provide us with the references of each type of radiator hoses and their mix with respect to the existing engines.

The references are like an engineering code that defines the degree of injection per second and the different mechanical dimensions. This information is strictly confidential. We were able to retrieve a table on the analysis of different types of motorization with their mixes in an anonymous way.

Quality Renault

The purchasing function is a transversal function. This is reflected in the RFI's drafting, which must specify the quality standards that Renault requires from its supplier.

To do this, the quality department recovers the quality objective of Renault for the years (2016, 2017, 2018, 2019) of the radiator pipes, as well as the degree of incident tolerated by quarter. This last criterion is measured by what is called the PPM indicator. This is means incident by thousands.

- For example: a PPM = 5 goal, means that in a number of 1000 parts delivered to the Renault site, the degree of scrap and non-conforming parts must not exceed 5.

Indeed, Renault is developing another quality indicator for its suppliers. This is the ranking. It classifies its suppliers and awards a quality rating based on the experiences with the supplier, as well as its reputation and ISO quality certifications

After the recovery of the quality objectives, it is necessary to do the same for the logistics side. Except that the objectives of the logistics are the same as the previous year with the precision of the incoterm FCA. This precision gives transport responsibility to Renault.

3.3.2 The technical file

The purchasing file consists of two main elements: the RFI and the technical file. The responsibility for the elaboration of this last one weighs on the shoulders of the engineers.

In fact, this file contains the technical definition of the parts, specifying the 2D and 3D models, and the technology used for quality production and consistent with Renault standards.

3.4 Analysis Phase

The following two indicators need to be analyzed:

- Transport encryptions.
- Current CA analysis.

3.4.1 Transport encryption

In order to calculate the transport gains that Renault can generate from this project, it is necessary to carry out a transport calculation, comparing the current logistic flow and the future flow of the radiator pipes. Below is a diagram of the two flows:

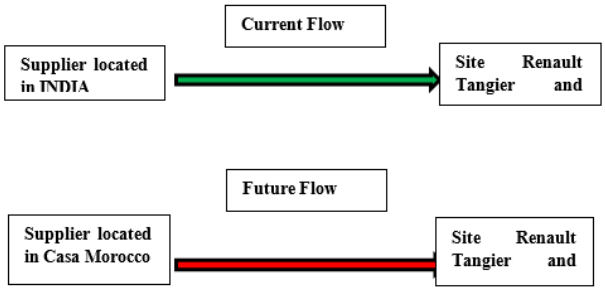


Figure 8: Logistic flows of radiator pipes

To compare the logistic costs of each stream there is a unit at Renault specializing in this transport encryption. It calculates the transport costs of the two flows.

The nature of the information collected for each reference is:

- The types of projects, in other words the different vehicles impacted.
- The weight of each reference: each reference has a specific weight. This information is provided by the engineering department of the company.

Daily rates: the volume that Renault wants to receive every day to avoid breaking the production chain.

- The code of the packaging unit (CPU): the type of CPU used in order to know the number of parts that this packaging unit can contain.
- The code of the handling unit (UM): the type of UM used in order to know the number of the CUs that the UM can contain.
- The supplier site account and the customer account.

After collecting this information from different departments, the launch of transport encryption for current flows will be possible.

After a kick off, the engineering department deducts the nature of the CUs and the UMs that the supplier must use since the transport is provided by Renault, consequently, Renault launched the transport costing for this future flow.

The return of the two transport figures, made it possible to identify the transport gain that Renault can earn.

What we have deducted is that the transport gain will be 5.51 euro for each family of parts delivered to the Renault plant.

For this purpose the Purchasing Department has established a current CA analysis and the prices that Renault pays for the radiator hoses in order to compare with the supplier's offer.

3.4.2 Current turnover analysis

The reference prices appear on a platform called SCOPP, which manages supplier contracts, their appointment, and the prices of all sourced references as well as supplier flows.

The list of references will be sent to a cell specializing in the administration of this platform, in order to retrieve the prices of each reference.

The objective was to realize the impact of total turnover of references consulted and to know the amount that Renault currently pays. Here is the result of the analysis in the form of a table:

Table 9: Turnover analysis

The parts/year	price
Radiator Inlet Hose	23 000,00 €
Radiator Outlet Hose	11 000,00 €
Heat Inlet Hose	35 000,00 €
Hose outlet radiator heating	38 000,00 €
Hose intake tank degassing	100 000,00 €
Hose outlet tank degassing	57 000,00 €
Hose purge	19 000,00 €
Total	**283 000,00 €**

The turnover shown in this table is the amount that Renault is paying today. Vis-à-vis its supplier, it is necessary to specify that this table summarizes the turnover for each type of radiator pipes.

Indeed, each type of radiator pipes corresponds to a sum of references. At the end we have specified in the project framework that there are 800 references for all radiator pipe parts.

3.5 Supplier return

The supplier return was set at the second week of April, unfortunately, Renault did not receive the supplier offer on time.

As a result, Renault relaunched its demand, and this time Renault received an unfavorable return which expresses that the quantity of references sent is very large and it will take a long time to encrypt them on the form of an industrial process. In addition it has even specified that to produce the quantity requested it will need very large investments which are very expensive and that the requested technology does not exist in Morocco. Therefore Renault must expatriate it from their site located in Ukraine.

In short, Renault has deduced that the supplier is not competitive on all references.

3.6 Presentation and interpretation of results

3.6.1 Presentation of the results

The SMA provider sent its quotation for the 20% references consulted. The supply generally consists of two parts. A part price of the piece and a logistics part. For the latter we have already carried out the transport costing and we have been able to show the transport gains. Here is the price suggested by SMA for the 12 references consulted (20% of total references):

Table 10: Offers SMA Casablanca

The references	Unit price of SMA in €
Reference 1	21,23 €
Reference 2	18,54 €
Reference 3	28,29 €
Reference 4	10,93 €
Reference 5	12,48 €
Reference 6	17,53 €
Reference 7	15,05 €
Reference 8	16,78 €
Reference 9	20,60 €
Reference 10	28,29 €
Reference 11	16,79 €
Reference 12	22,39 €
Total	**228,9 €**

The next step is to compare these prices with the current price. However, it must be taken into consideration that the encryption we have done previously was on all the parts, and consequently on all the references. For that we have redrafted our encryption to have our gain of transport only on the 12 referens concerned.

To summarize, we must compare the total supply. In other words the logistics part and the parts prices. Below the tables that presents the SMA offer and the offer of our current supplier with a precise and profound comparison:

Table 11 : Offers SMA Casablanca

	Offers SMA Casablanca		
The references	Unit price per piece	Current transport flows	Total price
Reference 1	21,10 €	0,13 €	21,23 €
Reference 2	18,35 €	0,19 €	18,54 €
Reference 3	27,89 €	0,40 €	28,29 €
Reference 4	10,66 €	0,27 €	10,93 €
Reference 5	12,10 €	0,38 €	12,48 €
Reference 6	17,20 €	0,33 €	17,53 €
Reference 7	14,23 €	0,80 €	15,05 €
Reference 8	16,20 €	0,58 €	16,78 €
Reference 9	20,10 €	0,50 €	20,60 €
Reference 10	28,12 €	0,17 €	28,29 €
Reference 11	16,54 €	0,25 €	16,79 €
Reference 12	22,13 €	0,26 €	22,39 €
Total	**224,62 €**	**5,26 €**	**228,9 €**

Table 12: The current offer of INAS India

	Current Supplier Offer INAS India		
The references	Current unit price of piece	Current transport flows	Total price
Reference 1	16,95 €	1,66 €	18,61 €
Reference 2	15,33 €	0,85 €	16,18 €
Reference 3	28,56 €	0,87 €	29,43 €
Reference 4	8,73 €	0,93 €	9,66€
Reference 5	8,66 €	0,79 €	9,45€
Reference 6	15,11 €	0,58 €	15,69 €
Reference 7	12,56 €	1,89 €	14,45 €
Reference 8	11,44 €	1,05 €	12,49 €
Reference 9	17,98 €	1,13 €	19,11 €
Reference 10	20,81 €	0,47 €	21,28 €
Reference 11	15,35 €	0,66 €	16,01 €
Reference 12	16,31 €	0,75 €	17,06 €
Total	**187,79 €**	**11,63 €**	**199,42 €**

And finally the table of the comparison of the two offers:

Table 13: Comparison of offers

Comparison of offers		
Price per piece	**Transport flow**	**Total Delta +/-**
4,15 €	-1,53 €	2,62 €
3,02 €	-0,66 €	2,36 €
-0,67 €	-0,47 €	1,14 €
1,93 €	-0,66 €	1,27 €
3,44 €	-0,41 €	3,03 €
2,09 €	-0,25 €	1,84 €
1,67 €	-1,09 €	0,58 €
4,76 €	-0,47 €	4,29 €
2.12 €	-0,63 €	1,49 €
7,31 €	-0,30 €	7,01 €
1,19 €	-0,41 €	0,78 €
5,82 €	-0,49 €	5,33 €
36,83 €	**-7,37 €**	**31,74 €**

The return of the SMA for the 12 references will allow us to check our local sourcing problem, and the results interpretation will be presented in the next chapter.

3.6.2 Results and interpretation

The competitiveness of suppliers is a fundamental criterion for supplier selection at Renault. Before the launch of the project we have established criteria that we want to find at the supplier consulted, namely competitive and profitability as well as performance.

The total price offered by SMA exceeds mostly the current price of INAS India even though we have gained in terms of transportation. But on the actual price of the piece the offer of INAS India is more competitive than that of SMA.

After an analysis to understand this price gap, we found that this difference is due to the large investments that SMA has set up to produce these references. A very advanced technology to produce only 12 references this increased the fixed costs and by default the cost price of SMA.

The lack of propulsion technology in Morocco was the main obstacle for this project. In addition there are other projects that are frozen for the same cause. When the supplier makes his calculations of the costs of installing the technology locally he finds that the project is not profitable and that his initial investment far exceeds the generated cash flows.

Indeed, most suppliers do not want to commit to low quantities especially for parts that require advanced technologies in their manufacturing process. The arrival of PSA in 2019 will make this demand more or less high and therefore will be an opportunity for Renault to unblock these frozen projects.

The levers and the stakes of local integration depend directly on the competitiveness of existing suppliers locally, so multinationals must seek to cut this dependence on suppliers in order to improve their performance through local sourcing.

It is obvious that the local integration makes it possible to gain in terms of logistics, but not in the case of parts prices. The installation of investments and the non-existing technology impacts the total price of the parts.

We decided to continue looking for another supplier to produce these parts, while Renault will continue to source from INAS India.

Our study led us to conclude that local sourcing does not always contribute to purchasing optimization and that sourcing to the International may be more competitive than local sourcing even if the company gains tremendously in terms of logistics.

Indeed, we have noticed that sourcing actually reduces logistics costs, but multinationals are looking to reduce their total purchase costs in terms of transport and in terms of products themselves.

4 General conclusion

Local sourcing is a very complicated process in non-industrialized countries. Thanks to a solid industrial fabric, an experience effect, and advanced technologies available within the country Renault was able to achieve a local integration rate of 90% in India.

The lack of technology and the effect of experience make it very difficult to establish local integration because investment costs are enormous. By default, the competitiveness of local suppliers in relation to international suppliers' prices will be impacted. Of course, ecosystem conventions have been established to solve this problem of non-competitiveness of local suppliers. This was possible through the subsidies that the Moroccan State undertakes to grant suppliers to encourage them to make the necessary investments.

Indeed, the problem of the departure was to know if the local sourcing always allows optimizing the purchase costs of the multinationals. To verify this hypothesis we have piloted a project of local integration for the multinational Renault.

We have found that local integration does not always optimize the purchasing costs of multinationals because of the lack of competitiveness of local suppliers, especially in developing countries.

BIBLIOGRAPHY AND WEBOGRAPHY

[1] Op.cit., Bozarth, Handfield, Introduction to Operations and Supply Chain management, Pearson, 2007

[2] Supply Chain Brain, U.S. and China Are the Riskiest Geographies for Sourcing and Manufacturing.

[3] AberdeenGroup, Low-Cost Country Sourcing : Success Strategies Maximizing and Sustaining the Next Big Supply Savings Opportunity, Juin 2005

[4] Sara Ireton, Global Sourcing Checklist. Industry Week, janvier 2007, p. 40

[5] Mike Jones, Domestic costs of foreign sourcing. Material Handling Management, février 2008, p. 40-41

[6] Sean Delaney, Delivering Value in both good and bad times, E-sourcing Forum, 14 juillet 2008

[7] .N. Sheth, « présentation d'un modèle de comportement des acheteurs industriels

[8] Le Supplier Choice Model » WOODSID & VYAS (1984;1986; 1987)

[9] Roger Pérotin marketing achats, stratégies et tactiques, Edition 2010

[10] SidhuSephardi, Sourcing strategy, principles, policy and designs, 2005, Springer

[11] Salons Solutions, discussion informelle avec des décideurs, CNIT Paris La Défense, 30 septembre 2008.

[12] Collectif, Encyclopédie de gestion, article « Stratégie d'entreprise et économie industrielle », écrit par A. Desreumaux, Economica, 1997.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Integration of the ASR Toolkit Kaldi into a Domoticz Home Automation System

Mohamed Aboulkhir[a], Samira Khoulji[b], Reda Jourani[c] and ML Kerkebd
Research Group "Information Systems Engineering", Abdelmalek Essaadi University, 93000,Tetouan,Morocco
[a,b,d,]National School of Applied Sciences, [c]Polydisciplinary Faculty of Tetouan,.
[a]mohamed.aboulkhir@gmail.com, [b]khouljisamira@gmail.com,[c]reda.jourani@yahoo.fr, kerkebml@gmail.com

ABSTRACT

This paper presents the design and the implementation of an interface between Kaldi, automatic speech recognition toolkit, and a home automation system. This interface is based on Open Platform communication (OPC) protocol. The developed architecture allows the injection of vocal data into a home automation system. Developed in C++, The Kaldi behaves as aclient of the OPC server.

Keywords : Speech recognition, domotics,Kaldi, Smart home, OPC.

1 Introduction

The technological evolution has made it possible to envisage many possibilities in order to improve the quality of life and to support people in their daily life. One solution to provide this assistance on a daily basis is the development of intelligent houses that are equipped with sensors, actuators, controllers and centralized software [1]. Various methods of interaction are being developed in this framework, but one of the most promising is the automatic speech recognition [2].

In our study, we will be interested to integrate Kaldi as an ASR system to home automation system. This integration will be based on OPC protocol. So we will develop a program on Kaldi that will allow him to behave like an OPC client in order to be able to finally integrate it into a home automation system. The expected result will allow a better interaction between the man and the intelligent house.

The paper is organized as follows: we start by describing the ASR software Kaldi. This is followed by describing components of home automation system and OPC protocol .Finally we will present our solution of integration of all these elements.

2 KALDI

Kaldi is an open-source toolkit for speech recognitionwritten in C++ and licensed under the Apache License v2.0[3].The architecture of Kaldi, as described in Figure 1, consists of the following modules:

- External Libraries: Kaldi depends on two external libraries that are also freely available. One is OpenFst for the finite-state framework (FST), and the other is numerical algebra libraries such as BLAS “Basic Linear Algebra Subroutines” and LAPACK “Linear Algebra PACKage”.

DOI: 10.14738/tmlai.54.3424
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3424

- Kaldi C++ Library: It contains all the functionalities and different modes of a speech recognition system developed by C ++, which are then called from a scripting language for building and running a speech recognizer.
- Kaldi C++ Executables
- Scripts

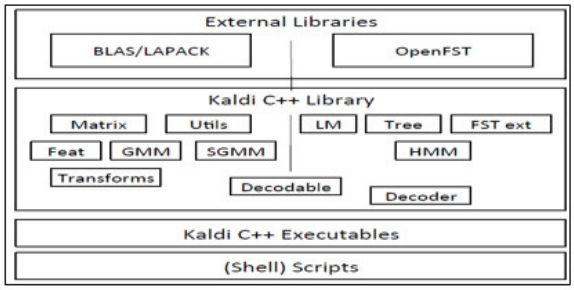


Figure. 1 – Kaldi architecture

The Kaldi library integrates all elements of an ASR system namely: Parameter extraction[4], acoustic model[5], language model[6].

2.1 Parameter extraxtion

Kaldi supports the most used parameters extraction methods. Thus, its Waveform Reading Code is intended to create standard MFCC and PLP functionality by setting reasonable defaults, but leaving available the options that people can modify. Kaldi supports the extraction approaches of most commonly used functionalities [7].

2.2 Acoustic modeling

Kaldi supports standard acoustic modelssuch as GMM, SGMM... and it easily extensible to new kinds of model.

Indeed Kaldi support GMMs with diagonal and full covariance structures. Therefore, it implements a GMM class that is parametrized by the natural parameters. The GMM classes also store the constant term in likelihood computation, which consist of all the terms that do not depend on the data vector. Such an implementation is suitable for efficient log-likelihood computation [8].

For subspace Gaussian mixture models (SGMMs), the toolkit provides an implementation of the approach in which all Hidden Markov Model (HMM) states share the same Gaussian Mixture Model (GMM)structure with the same number of Gaussians in each state[9].

2.3 Language modeling

As described in Kaldi's architecture, it uses a framework based on FST. It is therefore possible to use any language model that can be represented as a FST. Kaldi also offers several toolkits to convert language models into FST [10]

3 Home Automation System

A smart home is a combination of several advanced technologies. It consists of incorporating very small chips, which possess wireless communication, perception and information processing capabilities, in everyday articles. The goal is to create a transparent computing environment for the inhabitant. To be

able to provide services, the system must be able to acquire, process, and transmit information at all times. It must also be able to understand the needs of the user and control the various equipment in a smart way, in order to make the environment more comfortable. Moreover, it must be able to reduce the energy consumption without influencing the habits of the inhabitant [11].

Despite the large number of variations of the concept of home automation systems, there are nevertheless common modules. Figure 2 shows the general organization of a home automation system and of the generic modules constituting it[12].

In recent years these home automation systems have experienced the introduction of speech recognition systems. Thus, several projects have emerged such as the COMPANIONS, Companion Able, ALADIN or PIPIN projects. However, there are still many questions about the real impact, from both the point of view of assistance and the cost of installation, despite real evaluations in test apartments with the PERS, DIRHA projects [13].

Figure. 2– Home automation system

The objective of our study is to exploit the robustness of the Kaldi system in the automatic recognition of speech in order to integrate it in a home automation system in order to promote its functionality. Our study will be divided into two parts: first, a feasibility study of the solution chosen then a test experiment of robustness of the solution.

4 KALDI Integration into A Domoticz Home Automation System

The aim of our work is to integrate our chosen speech recognition software Kaldi, to a home automation system. For this, we will first propose an architecture of integration, which will be the object of several tests in order to qualify the performances of our model.

4.1 Integration Architecture

In a first step, we propose a communication architecture between Kaldi and a home automation system. Our proposal was based on an architecture based on the OPC client / server technique (Figure 3)[14]. This choice is motivated by:

- The C ++ language supports OPC. Thus, Kaldi can be configured as an OPC client [15].

- The client / server communication can be integrated with the most home automation systems[16].

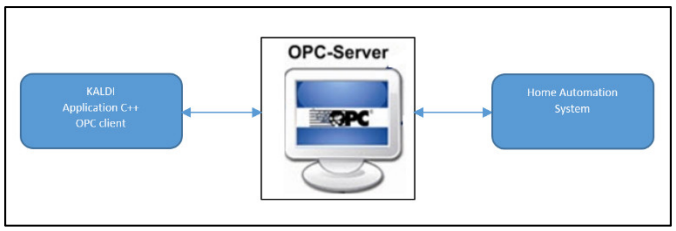


Figure. 3 Integration architecture

4.2 OPC

Open Platform Communication (OPC) is a technique developed in 1995 for the interoperability of industrial systems. OPC was designed to connect Windows applications and hardware and process control software.OPC servers provide a method that allows different software to access data from process control devices, such as a controller. Traditionally, whenever a program required access to a device's data, a custom interface, a driver, had to be written. The objective of the OPC is to define a common interface written once and then reused by any enterprise software. Once an OPC server is written for a particular device, it can be reused by any application capable of acting as an OPC client. An OPC server uses the Microsoft OLE (Component Object Model) technique to communicate with clients[17].

Today, OPC is a registered trademark of the OPC Foundation. The OPC specifications can be divided into two categories:COM / DCOM and Web Services. The first category includes:

- OPC Common (a common specification for all servers)
- OPC Data Access (access to real-time data)
- OPC Alarm and Event
- OPC Historical Data Access
- OPC Batch (batch processing)

The second category consists of a single specification divided into several parts OPC Unified Architecture.

The Unified OPC (UA) architecture has been defined and can be implemented with Java, Microsoft .NET or C, eliminating the need to use a Microsoft Windows computer with earlier versions of OPC. The UA combines the functionality of existing OPC interfaces with new techniques such as XML and Web services [18].

4.3 Feasibility study

In this first phase, we selected a home automation software called Domoticz, it is an open source Home Automation System. The experiment therefore consists in implementing the architecture described above and which is based on the OPC protocol.

4.3.1 Home automation system: Domoticz

Domoticz is an open source and free home automation software, consuming few system resources. It works on several OS (Linux, Windows). It allows to manage switches, detectors (fire, opening,

movement,), sensors (temperature, humidity, winds…), IP camera and to create virtual devices. All the configuration and use is done via a rather successful interface.

It is multi protocols Z-wave, RFX, En-Ocean, ... and it is possible to set up Lua scripts. It integrates with several systems of notification and sending of mail. The access can be secured via login / password with different level of authorization (spectator, user, and administrator). You can also allow only some IPs to connect.

4.3.2 Experiment

Here we report some outcomes about the feasibility of the interconnection between Kaldi and Domoticz through a server OPC. FIG. 4 shows the architecture adopted.

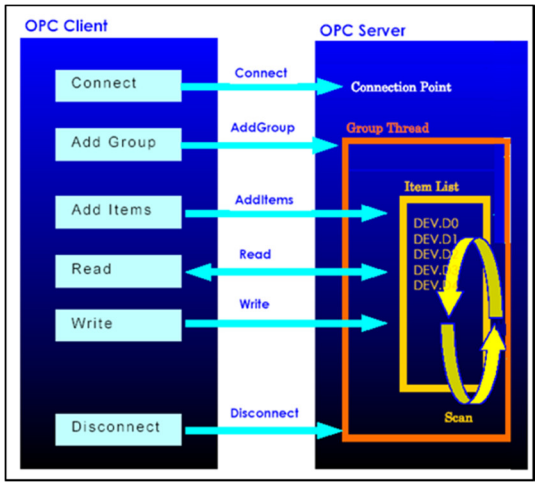


Figure. 4– OPC Architecture

Firstly, it is a question of developing scripts under C ++ to allow Kaldi and Domoticz to become OPC clients, and then the two systems will be connected to an OPC server which will allow d To use the voice data for controlling an output of the home automation system.

For example, we have tested the extinction of light by Domoticz (test at the output) if we pass through Kaldi the phrase "turn off the light".

Through these tests we have proved the feasibility of the interconnection between an ASR system and a home automation system through the developed architecture.

5 Conclusion

We described the architecture, which will allow the integration of a speech recognition system into a home automation system. This architecture was the subject of a feasibility study that showed the possibility of this integration. Development of this solution is continuing and we areworking on testing the robustness of our solution and studying the possibility of being generalized to all systems of home automation and ASR.

REFERENCES

[1] Noury, N., Virone, G., Barralon, P., Ye, J., Rialle, V., & Demongeot, J. (2003, June). New trends in health smart homes. In enterprise networking and computing in healthcare industry, 2003. Healthcom 2003.

[2] Frederic Aman. Reconnaissance automatique de la parole de personnes agées pour les services d'assistance à domicile. Traitement du signal et de l'image. Université Grenoble Alpes, 2014.

[3] The Kaldi Speech Recognition Toolkit, Povey Daniel, Ghoshal Arnab, Boulianne, GillesBurget, LukasGlembek, OndrejGoel, Nagendra, Hannemann, MirkoMotlicek Petr, Qian Yanmin, Schwarz Petr, Silovsky Jan, Stemmer Georg and Vesely Karel, Idiap-RR-04-2012

[4] Richard Dufour. Transcription automatique de la parole spontanée. Informatique [cs]. Université du Maine, 2010.

[5] Mohamed Bouallegue. L'analyse factorielle pour la modélisation acoustique des systèmes dereconnaissance de la parole. Autre [cs.OH]. Université d'Avignon, 2013.

[6] Insect sound recognition based on mfcc and pnn. In Multimedia and Signal Processing (CMSP), 2011 International Conference IEEE

[7] Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., ... & Woelfel, J. (2004). Sphinx-4: A flexible open source framework for speech recognition

[8] Sakriani Sakti, Oyunchimeg Shagdar, Fawzi Nashashibi, Satoshi Nakamura. Context Awarenessand Priority Control for ITS based on Automatic Speech Recognition. International conference on ITS Telecommunications, Dec 2015, Copenhagen, Denmark. 2015,

[9] D. Povey, L. Burget et al., "The subspace Gaussian mixture model A structured model for speech recognition," Computer Speech & Language,vol. 25, no. 2, pp. 404–439, April 2011.

[10] Fethi Bougares. Attelage de systèmes de transcription automatique de la parole. Ordinateur et societé [cs.CY]. Université du Maine, 2012.

[11] Michel Vacher, Benjamin Lecouteux, Javier Serrano-Romero, Moez Ajili, Fran_cois Portet, etal.. Speech and Speaker Recognition for Home Automation: Preliminary Results. 8th International Conference Speech Technology and Human-Computer Dialogue "SpeD 2015", Oct 2015.

[12] Augusto, J. C., Liu, J., McCullagh, P., Wang, H., & Yang, J. B. (2008). Management of uncertainty and spatio-temporal aspects for monitoring and diagnosis in a smart home. International Journal of Computational Intelligence Systems.

[13] Vacher, J.-F Serignat, Pelayo Menendez-Garcia, D Istrate. FIRST IMPLEMENTATION OF A SOUND/SPEECH REMOTE MONITORING REAL-TIME SYSTEM FOR HOME HEALTHCARE. The 6th International Conference Communications, Jun 2006

[14] Olivier Passalacqua, Eric Benoit, Marc-Philippe Huget, Patrice Moreaux. INTEGRATING OPC DATA INTO GSN INFRASTRUCTURES. IADIS International Conference APPLIED COMPUTING 2008

[15] Zheng, L., & Nakagawa, H. (2002, August). OPC (OLE for process control) specification and its developments. In SICE 2002. Proceedings of the 41st SICE Annual Conference (Vol. 2, pp. 917-920). IEEE.

[16] Topalis, E., Orphanos, G., Koubias, S., & Papadopoulos, G. (2000). A generic network management architecture targeted to support home automation networks and home internet connectivity. IEEE Transactions on Consumer Electronics, 46(1), 44-51.

[17] Liu, J., Lim, K. W., Ho, W. K., Tan, K. C., Tay, A., & Srinivasan, R. (2005). Using the OPC standard for real-time process monitoring and control. IEEE software, 22(6), 54-59.

[18] Grossmann, D., Bender, K., & Danzer, B. (2008, August). OPC UA based field device integration. In SICE Annual Conference, 2008 (pp. 933-938). IEEE.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Approaching Mental Disorders from the Engineering Point of View

J. Aguiló
Department of Microelectronics and Electronic Systems, Universidad Autónoma de Barcelona, Barcelona, Spain
Biomedical Research Networking Center (CIBER-BBN), Spain
acmlis.conference@gmail.com

ABSTRACT

Mental illness and mental disorders represent an increasing burden affectingthepopulation of all ages at all places, challenging mental health and health systems and contributing to the onset or to the acceleration of many other diseases. Cardiovascular risk, diabetes or also depression among many others seems to be clear examples of these effects. Concerning depression, there is a critical cross feedback: Stress causes depression whilethat staying in depressive states causes stress.

An accurate, reliable and almost continuous monitoring of the "instantaneous" state of the subjects would undoubtedly contribute to a better diagnostic and follow-up. So, treatments would be personalized by fine tuning relationships between drugs or other kind of interventions and subject state which in turn would contribute to a better knowledge of the causes of the disease and of the mechanisms acting in its treatment.

Following this primary thought,a research line explicitly focusing quantitative assessment methods and related technologies was stablished at the Universidad Autónoma de Barcelona.

First results had come from the study of moderate stress. In collaboration with various technological and clinical groups we proposed a reference scale to measure the stress level and we have taken the first steps concerning its validation; we are proposing a set of weighed parameters coming from physiological signals as a multivariable biomarker including some unexpected components. This "biomarker" show a strong correlation with the reference scale then stress can be quantitatively assessed continuously, reliably, repeatable and avoiding most of the subjective components of the patient and the observer.From these results, it is expected to demonstrate that a tool is available to determine, for example, the quality of an empowerment session, the effectiveness of an intervention using mindfulness in children with TDAH, or to determine the extent to which biofeedback techniques are useful as tools to improve quality of life of primary caregivers of chronic patients.

Some of the pilots planned at the very beginning are still underway while new research in Depression, Epilepsy, MS and Alzheimer is under way following similar guidelines.

Keywords-Mental disorders, Mental Health, Reference Scale of severity, Multivariable biomarker

DOI: 10.14738/tmlai.54.3651
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3651

1 Introduction

It is well known that from ancient Egypt, and probably before, the study of mental illness was of concern to society and to those in charge of science and health development, that even proposed and tried out surgical solutions to certain problems.

However, up to now advances in this area are not so fast than in most other medical fields. This is probably due to its inherent complexity and to the difficulties in finding atrue gold standard for testing methodologies allowing a reliable assessment.

Meanwhile, the numbers grow dramatically. The population in Europe is aging at an unprecedented proportion as the result of declining reproduction rates and increasing life expectancy. Because of that, chronic diseases are the main cause of illness in old age but mental disorders such as dementia and late-life depression play also a significant role as well as epilepsy and Alzheimer. Besides, because of the constant pressureof the modern way of life imposes on the individuals, stress is also dramatically growing-up to point that the World Health Organization called it a World-Wide Epidemic.

In particular, the number of patients with dementia is expected to rise sharply as the prevalence of dementia doubles every 5.2 years exponentially between 65 and 85 years of age. Late-life depression is also an important public health problem. Estimated 1-year prevalence rates of depression range from 3% to 10%. For dementia, prevalence rates are estimated between 0.6% and 3.7% for 65-year-old to 69-year-old, and 25.2% to 75% for adults 90 years old or older. Additionally, multiple studies have shown high comorbidity between mental disorders and chronic physical illnesses in the elderly. Positive feedback has also been demonstrated between different diseases such as diabetes mellitus, cancer and cardiovascular disease and the occurrence of depression in the elderly. And vice-versa, in older populations, there is also a positive feedback between depression and hypertension, diabetes and cardiovascular illnesses.

As it was pointedout by the Global Alliance for Chronic Diseases, “Mental disorders represent an ever-increasing burden, to all ages of the population, challenging mental health and health systems. For example, depression affects 350 million people in communities across the world and represents the third leading contributor to the global disease burden. Dementia affects 47.5 million people worldwide with 58% of people living with dementia in low- and middle-income countries. Global costs associated with mental disorders were estimated to be $2.4 trillion in 2010 and are expected to rise to $5.8 trillion by 2030. Mental disorders cause tremendous suffering for individuals, families, communities and societies. They also increase the risk of co-morbid illnesses and social exclusion. There are obstacles to achieving effective prevention, early identification and management of mental disorders and to ensuring patients' adherence to therapies.

Coming from the Micro-Nano-Bio Systems design and prototyping to solve problems in graft transplant, in heart surgery, in functional recovery of ear and vision or in early detection of glaucoma, the GBIO research group from CIBER-BBN decided to explore possible technology help in monitoring mental health and wellbeing.

Two amazing things comes up from the first moment to the engineering-based team facing the quantitative assessment of emotional or psychological stress: the lack of a precise definition except in very severe manifestations and the lack of a reliable, totally objective and repeatable method of the severity measurement method. These two things were the first topics to afford.

Although we are now working with different emphasis in Depression, TDAH, Alzheimer, Multiple Sclerosis and Epilepsy we start by trying to study stress. We thought that stress, when not severe or traumatic stress, were not a real illness so, in this way, we would not interfere with medical research. Besides, to implement results that we would have, if any, we will need to assume responsibility only concerning individual wellbeing of treated individuals, not concerning its health.

Once focus on moderate stress was decided, the main aim of the study would be summarized as attempting to quantify the level of stress as well as to find and to develop appropriated devices and/or methodologies to have an objective, reliable, and repetitive measurement procedure set-up.

Availability of such a tool will facilitate precise communications among professionals, allow a close follow-up of patients, will help doctors in diagnostic but also in the prescription of the medication and in quantifying the efficacy of the drugs. Establishment of thresholds will become also available in such a way that illness condition or not would be quantitatively established as well as related risks in some professional situations such as drivers, air controllers, police, firefighters, army, etc.

2 Methodology and Experiment's Organization

As already stated, the first step was to establish a definition of stress that would be acceptable from thepsychiatric point of view whilst clear from the engineering point of view. We took the following statement as stress definition: The "distance" measuringhow close or how far is the individual positioned from the medically and socially considered the "normal" status. This distance measured in a n-dimensional space where all the parameters related to stress should be considered. When considering acute stress this distance is considered between the everyday typical values of all considered variables, that is its point of homeostatic equilibrium, and the actual values of them after the stress stimulus. In chronic stress, the "normal" status should be taken among all the population with similar characteristics.

Concerning stress related information, first of all, we took as reference the psychometric test commonly used by professionals together with a commonly used and well documented moderate stressor. Stressor stimulus can be induced in the laboratory through a mental arithmetic task, public speaking task, by color word test, by hyperventilation, seeing film fragments, or using cold pressor test, among. Combining some of the already mentioned stressors, the Trier Social Stress Test (TSST) is so far one of the widely used and documented protocols in stress research [1], and it has been shown to exert major effects on almost all psycho-biological response mechanisms [2]. In a similar way, the autogenic relaxation method was used to induce a relaxed stated that will be used as the basal state of each individual.

On the other hand, the most frequently used psychometric test concerning stress are: the subject's self-report assessments, the Visual Analogue Scale for Stress (VASS), the State-Trait Anxiety Inventory test (STAI-s). We also include some tests to evaluate individual trait against a stress stimulus. We take the (STAI-t) and the Perceived Stress Scale(PSS)

Researchers studying animal response to stress stimulus are usually using a set of specific biochemical markers to evaluate themagnitude of the response. Mainly they are using catecholamine, prolactin, copeptin, glucose, cortisol, α-amylase and recently some inflammation markers such as interleukin-6 and/or 10 or other cytokines as well as some tumor markers such as the TNF-a. Following this way, they have a more or less complete information about the endocrine and nervous responses of the organism to the stress stimulus.

Nevertheless, biochemical markers are an invasive measure, expensive and its correct collection depends on the time gap since the stressor is active. On the other side, psychometric tests are very related with individual subjectivity and lose reliability when they are frequently applied.

It is clear, for example, that emotions, stress and some mental disorders are altering the cardiovascular function through complex interactions that include feedforward and feedback mechanisms involving effects of the autonomic nervous system (ANS) on it. Processing electrocardiography (ECG) signals to extract further information usually consists of analyzing time series representing the heart rate, intervals related to the activation and repolarization of the myocardium, systolic and diastolic blood pressure, and markers related to pulse wave velocity and peripheral resistance. Then respiration and chest movement, mechanical properties of the vasculature, such as the blood pressure and volumetric measures of the vessels seems to be direct or indirectly related to the stress related changes in the homeostatic equilibrium. A similar argument could be applied to electromyography (EMG) and also to skin temperature and skin conductance. All these signals could become key information sources that in fact are already used to evaluate stress effects in specific conditions.

Table 1. Experimental pilots set-up

Type	***Title and short description***	***Responsible Institution &Place***
Psychological Stress	Relax vs. stress induced by the TSST in Healthy young subjects	U. Aut. Barcelona U. Poltech. Madrid U. Zaragoza
	Parents of children diagnosed with a severe surgery	H. St Joan de Deu
	Patients diagnosed of Depression syndrome	H.C. Zaragoza H. St Joan de Deu
Chronic stress	Caregivers of chronic patients	H. C. Barcelona
	Caregivers of terminal oncologic	H. C. Barcelona
Traumatic stress	Knee arthroplasty	H. C. Barcelona
	scoliosis surgery in children	H. St Joan de Deu
	Craniostenosis surgery in children	H. St Joan de Deu
Thermal stress	Healthy young subjects 6 h. at 35 ºC, 50% humidity	U. Aut. Barcelona
Asthma vs Psychological stress	Asthma patients	H. St Pau, Barcelona
Effect of mindfulness interventions in TDAH	Children diagnosed of TDAH syndrome	H. St Joan de Deu
The role of physical and cognitive information in DCL	Recent diagnosed Alzheimer patients	Fund. ACE Barcelona
Effect of biofeedback interventions in caregivers	Caregivers of terminal oncologic patients	H. C. Barcelona

In a mono-variate approach, these methods can be used to analyze the variability of single time series. In particular, the heart rate variability has been extensively studied because of its close interaction with the activity of the sympathetic and parasympathetic nervous systems. The spectral or time-frequency representation of the heart rate variability shows two main oscillations (called low frequency andhigh frequencycomponents) whose ratio can be used to monitor the sympathovagal balance.

The multivariate analysis of the interactions between cardiovascular and cardiorespiratory signals gives further information about the ANS modulation. Novel frameworks have been recently proposed to describe the dynamic (causal) interactions between cardiovascular variables, which include the strength of the coupling, the direction of information flow and the latency in coupled changes. For example, mental stress and emotions have been shown to induce consistent pattern of responses in these interactions, which can be used to continuously monitoring the psychophysiological state in patient follow-up.

To cover the different types of stress, while attemptingto gain information that can be useful regardlessof the stressor, the methodis structured around nine initial pilot studies conducted indifferent samples of individuals that involve differenttypes of stress, as well as varying types and intensitiesof stressors or application times. Each of the pilotsis a prospective observational study with a specific protocol that wasapproved by the correspondingEthics Committee. A priori estimations indicatedthat a sample that could be used as a statistically significant indicator for each individual trial and thatwas powerful enough to contribute to thedevelopment of the joint algorithm must consist ofno fewer than 30 individuals. Forty volunteer subjectswere selected for each trial, all of whom fulfilledthe required eligibility criteria, in the hope that theminimum figure would be reached despite the dropoutsand those excluded due to errors [3].

Four additional experimental pilots have been added during the last two years due to the research interests of the hospital medical community.

Each trial consists of studying the action of aparticular stressor, which could be real or inducedartificially. In each of these sessions, which correspondto different states of the subject, the actionof the stressor upon the electrophysiological, biochemicaland psychometric variables established inthe corresponding protocol was evaluated while atthe same time the temporal sequences of their valueswere also collected [4].

In most of the trials the reference was taken as beingthe actual individual under conditions that couldbe considered his or her normal state. In the trials inwhich this is not possible, the reference used was asubject with the same characteristics, paired by age,sex, etc., but without the presence of the stressor.

The method implemented attempts to avoid theerrors and uncertainty due to the type or form ofmeasurement. Hence, homogeneous protocols weredesigned for the taking of measurements and processing of samples for the different trials. The electrophysiologicalmeasurements are performed usingthe same calibrated commercial equipment with CE certification following a single procedure for placingthe sensors and for recording. Each of the biochemicalvariables (analytes) are determined in the samespecialized laboratory that will process the samplesfrom all the pilots that include the valuation of this analyte. The same considerations were applied to the analyses to be carried out on saliva and hair.

The recent technological effort in the biomedical field is including the incorporation of sophisticated but affordable electronic sensors, but also means a dramatic increase in calculating capacity and thus the use of sophisticated algorithms allowing to extract more and more information from simple and apparently well know signals.

An example already well developed is the "fine" analysis of the ECG signal as mentioned above. The ECG signal is affected by the electronic noise of the environment, but also by breathing, movement or muscle contractions, speech and even by the ambient or body temperature among others. A careful selection of

time windows with little interferencestogether with sophisticated signal processing algorithms to correct the aforementioned effects,is contributing to extract a greater amount of information from ECG and thus a greater and better knowledge of the behavior of the cardiovascular system but also of the whole organism including nervous system and brain functions. Our group recently contributed to this aspect by including the respiratory frequency information in the Heart Rate Variability, HRV, for a better analysis of Stress [5].

Following and extending this way, infrared cameras can be used in the near future, for the contact-less and remote record of the ECG and PPG signals, the skin temperature at different points and possibly the skin conductance. And from that the Pulse Transit Time, PTT, and the arterial pressure will be easily derived. And even more so, in a few years the voice signal recorded and processed from our own smartphone will be enough to predict a heart attack or an epilepsy seizure and warn the doctor and / or family if a memory failure that can cause problems is occurring.

Table 2. Set of parameters extracted from ECG, PPG and ST

Indices from HRV	*Indices from PPG*	*Indices from skin T*
HR_{mean} (bpm)	PR_{mean}	ΔT^{FI}
SDNN (s)	PP_{SDNN}	
RMSSD (s)	PP_{rMSSD}	$T_{fi_{Ptotal}}$
PPNN (%)	PDT_{mean}	
Power VLF (s-2)	PDT_{STD}	ΔT^{Fa}
Power LF(s-2)	PRT_{STD}	
Power HF(s-2)	PTT_{mean}	$T_{fa_{Ptotal}}$
Power PLF(s-2)	PTT_{STD}	
Power PHF(s-2)	PWr_{STD}	T_{ratio}
Total power(s-2)	-	
LF/Hf ratio	-	

Methods and algorithms to extract information from this sort of multimodal databaseare diverse too, from simple statistics to apply sophisticated classification techniques before using machine learning algorithms such as support vector machines, fuzzy logic, Bayesian networks, classification trees, artificial neural networks, Computer Aided Diagnostic tools orMulti-Task Gaussian processes.

Clearly, all these technological contributions are paving the road to a new paradigm in health care. The way in which data with diagnostic value are collected, analyzed and stored will change radically. The amount of available data will increase dramatically, offering the opportunity to evaluate diagnostic tools and treatments in a more accurate way and by using completely different equipmentSufficiently friendly and robust to be used mostly by the user itself.

3 Why A Reference Scale?

Measuring is a basic process carefully used in both science and engineering. It consists of comparing with a known pattern the phenomenon whose magnitude is desired to value. In many cases, mental health clinics lack this kind of “pattern”, because psychometric test could be easily falsified by the patient and cannot be applied too frequently.

Nevertheless, measuring is crucial to communicate to others, to advance knowledge rapidly and to facilitate comparison. Medicine is no stranger to it and, with increasing force, tries to establish principles and establish solid foundations and links between diagnosis and follow-up based on objective measures.

To measure it is necessary to have a reference: A comparison pattern allowing to quantitatively assess the magnitude to be measured. Conscious of the need, it was decided to take the pattern followed in 1985 at the George Washington University Hospital, Washington D.C., to score severity-of-disease of patients arriving at the Intensive Care Unit in such a way that higher scores correspond to more severe disease and a higher risk of death [6][7]. Physical, Psychological and Cognitive aspects were taken into account by using psychometric test and biochemical parameters obtained from experiments, for a primary components statistical analysis. As a result, the first approach to an acute stress reference scale has been proposed [8]. Proposed scale has been validated and proven to be robust measuring moderate psychological stress in healthy youngsters. Nevertheless, it is clear that proposed scale can't be widely used due to its inherent invasivity and high cost. To avoid these problems, we are elaborating a "working-copy" of this reference scale based on electrophysiological signals.

4 Expected and Unexpected Preliminary Results

The recent experimental pilots already analyzedare providing data and specific information, some of them surprising,that will be used in the design of new experiments and in the design of new hardware and software tools to help in diagnostic and follow-up.

Some of them are the following:

A. A reference scale to quantitatively assess the stress level has been proposed and validated.

From the engineering point of view, this is a very important result that will allow faster advances based on the scientific communication using a reliable and repeatable measurement tool.

B. Based on electrophysiological signals we have now a stress's multivariable biomarker

Using a well-known set of physiological signals, the level of stress of an individual with a statistically significant correlation with biochemical markers and/or psychometric tests can be determined.

Using this multivariate biomarker, it is possible to discriminate between 5 stress levels even within the moderate stress category resulting from the use of TSST.

C. Skin temperature seems to perfectly follow changes in the sympathetic-parasympathetic balance.

Results found on the analysis of moderate stress in healthy young individuals show that temperature is changing in agreement with stressor stimulus with minimal delay times. Following this line, an infrared digital camera or a normal digital camera with appropriate filters could constitute a totally non-invasive first step in the stress analysis.

D. Glucose concentration in blood as well as a-amylaseand probablyglycatedhemoglobin could be taken as bio-markers of the chronic stress in adults.

Since for the measurement of blood glucose there are many devices today, this can be a chronic stress biomarker of widespreaduse.

An explanation could be found for evidence based on statistical studies that stress induces diabetes.

E. It is not demonstrated that Cytokines such as IL-6 and IL-10 are good biomarkers for chronic stress.

Although some studies indicated that inflammation markers such as IL-6 and IL-10 may be markers of chronic stress, it has not been demonstrated in this study. Again, there are indications that point in this direction without being able to be taken as definitive.

F. The effectiveness of treatments can be measured.

The door to a tool to assess the quality of interventions using techniques of empowerment, mindfulness or biofeedback, drug based interventions or other forms of treatment is now opened.

5 Related Challenges to Face

A. Fine-tuning the reference scale

Only one stress scale? Only a single type of disorder under the stress name?

B. Quantitative assessment of stress: a quality measurement of empowerment sessions.

Results had and a proposal on what to do

C. Applying similar methodology to other mental disorders

Depression, Multiple Sclerosis, Epilepsy, Alzheimer, TDAH

D. Using developed scale and developed tools

The behavior under stress conditions as a biomarker for some disorders

E. Converging Technologies to remove the border along psychological and somatic responses?

Demonstrating quantitative associations between stress and diabetes, ictus, and heart attack between others

Could Technology and methodology set-up help in the early detection of anomalous behavior in young (including suicide) ?

ACKNOWLEDGEMENTS

This research line is supported by the CIBER-BBN(http://www.ciber-bbn.es)and for two of its consortium institutions: the Univ. Autónoma de Barcelona and the CSIC through the IMB-CNM (www.imb-cnm.csic.es/). Direct funding is coming from MINECO (FIS-PI12/00514 and TIN2014-53567-R) and from IMI (https://www.imi.europa.eu/) which is public-private initiative under the framework of the Horizon 2020 the European research and innovation program.

REFERENCES

[1] B. M. Kudielka, D. H. Hellhammer, and C. Kirschbaum, “Ten years of research with the Trier Social Stress Test - revisited,” in Social Neuroscience, 2007, pp. 56–83.

[2] K. L. Tamashiro, R. R. Sakai, C. A. Shively, I. N. Karatsoreos, and L. P. Reagan, “Chronic stress, metabolism, and metabolic syndrome.,” Stress, vol. 14, no. 5, pp. 468–474, 2011.

[3] Aguilo J, Ferrer-Salvans P, Garcia-Rozo A, et al. Project ES3: attempting to quantify and measure the level of stress. Rev Neurol. 2015;61(9):405-415.

[4] Arza A, Garzón JM, Hemando A, Aguiló J, Bailon R. Towards an objective measurement of emotional stress : Preliminary analysis based on heart rate variability. In: Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. Milan, Italia, Agosto 2015.

[5] Hernando A, Lazaro J, Gil E, Arza A, Garzón-Rey JM et al. Inclusion of respiratory frequency information in Heart Rate Variability for a better analysis of Stress.IEEE JBHI2016; 20 (4): 1016-1025

[6] Knaus W, Zimmerman J. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Crit care. 1981;9(8):591-7.

[7] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818-829.

[8] Jorge M. Garzón-Rey, Adriana Arza, Conchita de la Camara, Antonio Lobo, Antonio Armario, Jordi Aguilo. An Approach to an Acute Emotional Stress Reference Scale

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Random Key Cuckoo Search for the Quadratic Assignment Problem

Aziz Ouaarab*‡, Bela¨ıd Ahiod*, Xin-She Yang†
*LRIT, Associated Unit to the CNRST(URAC) no 29, Mohammed V-Agdal University, Morocco
†School of Science and Technology, Middlesex University, The burroughs, London NW4 4BT, UK
‡School of Technology Essaouira, Cadi Ayyad University. Morocco
aziz.ouaarab@gmail.com; Ahiod@fsr.ac.ma; x.yang@mdx.ac.uk

ABSTRACT

This paper proposes an adaptation of the Random- Key Cuckoo Search (RKCS) algorithm for solving the famous Quadratic Assignment Problem (QAP). We used a simplified and efficient random-key encoding scheme to convert a continous space (real numbers) into a combinatorial space. We also considered the displacement of a solution in both spaces by using Le´vy flights. The performance of the RKCS for QAP is tested against a set of benchmarks of QAP from the well-known QAPLIB library, and the comparison with a set of other methaheuristics is also carried out.

Index Terms—Nature-Inspired Metaheuristic, Cuckoo Search, Le´vy Flights, Combinatorial Optimization, Quadratic Assign- ment Problem, Random-Key.

1 Introduction

NP-hard problems [17] are very challenging to solve. It is also the most complicated among combinatorial optimization problems. The main difficult of such problems is that the number of combinations grows exponentially with the problem size. Quadratic assignment problem [7] is one of the problems that belongs to this class.

Quadratic Assignment Problem (QAP) is a combinatorial optimization problem that is applied to solve various problems in many fields such as Steinberg Wiring Problem [5], Hospital Layout [13], Dartboard Design [12], and many other applica- tions [7], [14]. Problems such as QAP do not have an efficient algorithm to solve them exactly. It is practically very difficult to get a solution of optimal quality and in a reduced runtime simultaneously. This requires some heuristic algorithms that can find good (not necessarily optimal) solutions in a good runtime by trial and error. Approximate algorithms such as metaheuristics [4] are actually the best choice to solve many combinatorial optimization problems. They are characterized by their simplicity and flexibility while demonstrating remark- able effectiveness. Many metaheuristics are proposed to solve QAP. Existing studies include Genetic Algorithms (GA) [1] that used a sequential constructive crossover, a modified Particle Swarm Optimisation (PSO) [21], Ant Colony Optimisation (ACO) [11], and many other examples [22], [15].

Among the most difficult issues that arises when solving a combinatorial optimization problem with a metaheuristic, is how to move in the combinatorial solution space without affecting the performance of

DOI: 10.14738/tmlai.54.3666
Publication Date: 28th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3666

the metaheuristic. Several meta- heuristics are designed in principle for continuous optimization problems. So, the question is how to treat combinatorial problems properly without losing the good performance of these metaheuristics. In this paper, we used the Random-Key Cuckoo Search (RKCS) algorithm by using the random key encoding scheme to represent a position, found by the cuckoo search algorithm, in the QAP solution space.

This work presents a novel approach using the improved cuckoo search algorithm [26], based on random keys [3], with a simple local search procedure to solve QAP.

The rest of this paper is organized as follows: Section 2 introduces briefly the QAP. Section 3, first, introduces the standard cuckoo search and its improved version. Section 4 presents the random-key encoding scheme, while Section 5 describes how CS solves QAP by using Random key. Then, Section 6 presents results of numerical experiments on a set of QAP benchmarks from the QAPLIB library [8]. Finally, Section 7 concludes with some discussions and future directions.

2 Quadratic Assignment Problem

The Quadratic Assignment Problem is a combinatorial optimisation problem, which tries to minimize the total cost of building and operating the facilities knowing that the benefit resulting from an economic activity at any site is depending on the sites of the other facilities. The solution space in QAP is considered as a set of all potential assignments of the facilities to the possible sites.

A selected solution S is the permutation φ of a given set Q = {1, 2, . . . , N } where N is the instance dimension, it is also the number of sites and facilities, φ(i) = k means that the facility i is assigned to the site k.

The objective problem is to find a permutation φ = (φ(1), φ(2), ..., φ(N)) that minimizes

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij} b_{\phi(i)\phi(j)} \qquad (1)$$

where a is the flow matrix, and aij is the flow between facilities i and j, and b denotes the distance matrix. So, the distance from facility i to j takes the value of bφ(i)φ(j).

Here, φ(i) is the location assigned to facility i. The aim is to minimize the sum of products flow×distance [30].

3 Cuckoo Search Algorithm

In the aim to increase their survival chances and reduce the probability of abandoning eggs by the host birds, cuckoos adopt many strategies and tricks. These strategies are mimicked successfully and designed in the well known Cuckoo Search (CS) algorithm [34]. Cuckoo search introduces Le´vy flights [28] for generating a new good solution. Le´vy flights, named after the French mathematician Paul Le´vy, represent a model of random walks characterized by their step lengths which obey a power-law distribution [6]. CS is widely applied to solve many combinatorial optimization problems such as Travelling Salesman Problem [25], [27], Flow Shop Schedul- ing Problem [20], Knapsack Problem [19].

The first version of CS, which is developed by Xin-She Yang and Suash Deb, is summarized as the following ideal rules: (1) Each cuckoo lays one egg at a time and selects a nest randomly; (2) The best nest with the

highest quality of egg can pass onto the new generations; (3) The number of host nests is fixed, and the egg laid by a cuckoo can be discovered by the host bird with a probability pa $\in$ [0, 1]. pa is also a switch parameter to control the balanced combination of local

explorative random walk and the global explorative random walk. The local random walk can be written as :

$$x_i^{t+1} = x_i^t + \alpha s \otimes H(p_a - \epsilon) \otimes (x_j^t - x_k^t), \qquad (2)$$

where x_j^t and x_k^t are two different solutions selected randomly by random permutation, H(u) is a Heaviside function, ϵ is a random number drawn from a uniform distribution, and s is the step size. On the other hand, the global random walk is carried out by using Lévy flights

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda), \qquad (3)$$

Where

$$L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi\lambda/2)}{\pi} \frac{1}{s^{1+\lambda}}, \quad (s \gg s_0 > 0). \qquad (4)$$

Here $\alpha > 0$ is the step size scaling factor, which should be related to the scales of the problem of interest. Lévy flights have an infinite variance with an infinite mean [34]. Here s_0 is a constant, which can be take as 0.01 to 0.1.

Before applying CS to solve QAP, as described in the Algorithm 1, we will consider an improved version of CS [26]. This improvement introduces a new category p_c of cuckoos that can engage a kind of surveillance on nests likely to be a host. So, around the pertinent solutions, this portion p_c intensify the search process to find a new better solution via Lévy flights.

Algorithm 1 Improved CS Algorithm

```
Objective function f(x), x = (x_1, ..., x_N)^T
Generate initial population of n host nests x_i (i = 1, ..., N)
while (t <MaxGen) or (stop criterion) do
    Start searching with a fraction (p_c) of smart cuckoos
    Get a cuckoo randomly by Lévy flights
    Evaluate its quality/fitness F_i
    Choose a nest among n (say, j) randomly
    if (F_i > F_j) then
        replace j by the new solution;
    end if
    A fraction (p_a) of worse nests are abandoned and new ones are built;
    Keep the best solutions (or nests with quality solutions);
    Rank the solutions and find the current best
end while
Postprocess results and visualization
```

4 Random-Key Encoding Scheme

Random-key encoding scheme [3], [27] is an interesting procedure that can be useful to pass from a continuous space to the combinatorial space. In general, the position in the continuous space is represented by a vector of real numbers. To have a projection in the combinatorial space, random-key (RK) associates each real number with a weight. These weights are used to generate one combination as a solution. The random real numbers drawn uniformly from $[0, 1)$ compose a vector showed in Figure 1. On the other hand, the combinatorial vector is composed of integers ordered according to the weights of real numbers in the first vector, illustrated in the Figure 1.

Random keys:	0.8	0.5	0.7	0.1	0.4	0.2
Decoded as:	6	4	5	1	3	2

Figure. 1. Random key encoding scheme

5 Random Key Cs For Qap

A QAP solution is a vector of N integers. Each integer is the facility index and its order in the vector is the corresponding site index. By considering Figure 1, we can say that resulting vector in this figure is a QAP solution. So to move from the current solution to a new one we can just perturb the first vector that contains the real numbers. This perturbation is performed via Lévy flights. In the case of big jumps we per- form a random pairwise interchanges of vector integers. It is a series of chained swaps $\phi_{\xi(1)\xi(2)}$, $\phi_{\xi(2)\xi(3)}, \ldots, \phi_{\xi(\rho-1)\xi(\rho)}$, where $\xi(1), \xi(2), \ldots, \xi(\rho)$ is a sequence of random integers between 1 and N such that $\xi(i) = \xi(\mathbf{j}), i, \mathbf{j} = 1, 2, \ldots, \rho, i = j, 1 < \rho \leq N$. The mutated solution π can thus be represented as a composition $((((\pi \oplus \phi_{\xi(1)\xi(2)}) \oplus \phi_{\xi(2)\xi(3)}) \oplus, \ldots) \oplus \xi(i)\xi(i+1)) \oplus \ldots) \oplus \phi_{\xi(\rho-1)\xi(\rho)}$, where π is the current solution. This series of chained swaps is called the controlled chained mutation (CCM) [23]. The parameter ρ is the mutation rate. This allows an improved way to balance the search for solutions in local areas as well as global areas.

To detect the best solution in the found area, a simple local search (Steepest Descent [2]) is performed. In this local search method we used 'swap' move as described in Figure 2 (In this example we chose to swap the facilities of the sites 2 and 6, which are 1 and 9.) which moves from a placement φ to a neighbour placement π by applying a swap between facilities r and s:

$$\pi(k) = \phi(k), \quad \forall k \neq r, s \tag{5}$$

$$\pi(r) = \phi(s), \quad \pi(s) = \phi(r) \tag{6}$$

To gain some precious time, estimations are generally performed. They restrict positively the choice of passing to a new good solution.

In the case of symmetrical matrices with a null diagonal the cost $\delta(\phi, r, s)$ of a move is given by:

$$\delta(\phi, r, s) = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(a_{ij}b_{\phi(i)\phi(j)} - a_{ij}b_{\pi(i)\pi(j)}\right) \quad (7)$$

$$= 2.\sum_{k\neq r,s}\left(a_{sk} - a_{rk}\right)\left(b_{\phi(s)\phi(k)} - b_{\phi(r)\phi(k)}\right) \quad (8)$$

In the minimization case, the swap move is done only if the new solution cost is less than the current one. Obviously, this process is repeated until no further improvement is possible or when a given number of steps is reached.

Steepest Descent is a simple local search method that can be easily trapped in a local minimum and, generally, it cannot find good quality solutions. We choose this simplified local search "Steepest Descent" method to show the performance of CS combined with RK for QAP. It allowed us to generate solutions of good quality, without introducing an advanced local search method.

6 Experimental Results

We will show some results of running RKCS to solve a set of bechmark instances [29], [31], [32], [10], [9], [13], [18], [33], [24] of QAP from the QAPLIB library [8]. Forty-six instances are considered with sizes ranging from 12 to 100 facilities. The numerical value in the name of an instance represents the number of provided facilities, e.g., the instance named sko90 has 90 facilities. We note that for each instance, 10 independent runs are carried out. These results describe the performance of this first version approach.

In Table II, we compare RKCS with two algorithms based on Genetic Algorithm [1] (SCX) and Ant Colonies [15] (HAS- QAP). Another comparisons are carried out to a Particle Swarm Optimization based algorithm [16] (HPSO) in Ta- ble III. We have implemented RKCS algorithm using Java under 32 bit MS Windows Seven operating system. Experi- ments are conducted on a laptop with Intel(R) Core™ 2 Duo 2.00 GHz CPU, and 3 GB of RAM. SCX algorithm has been encoded in Visual C++ and run on a PC with Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz and 8.00GB RAM under MS Windows Seven. For HAS-QAP it was not possible to obtain the hardware configurations used in the experiments. HPSO has been run on Intel pentium core 2 Duo Q9950 (2.83GHz).

TABLE I
PARAMETER SETTINGS FOR RKCS ALGORITHM.

Parameter	Value	Meaning
n	20	Population size
p_c	0.6	Portion of smart cuckoos
p_a	0.2	Portion of bad solutions
$MaxGen$	500	Maximum number of iterations
α	0.01	Step size
λ	1	Index
ρ	$Levy$	Mutation rate

The properly selected parameter values used for the experiments of RKCS algorithm are shown in Table I. These values are selected, based on some preliminary trials, and gave the best results concerning both the solution quality and the computational time. In each case study, 10 independent runs of RKCS with these parameters are carried out.

TABLE II
COMPARING RKCS WITH GENETIC ALGORITHM (SCX) [1] AND ANT COLONIES (HAS-QAP) [15].

		SCX		HAS-QAP		RKCS	
Instance	Bkv	PDbest(%)	time(s)	PDbest(%)	time(s)	PDbest(%)	time(s)
tai20a	703482	4.50	6	1.483	3	**0.0**	20
tai20b	122455319	6.56	6	0.243	3	**0.0**	0.5
tai25a	1167256	4.58	9	2.527	5	**0.0**	39
tai25b	344355646	5.24	9	0.133	5	**0.0**	1
tai30a	1818146	4.29	13	2.600	9	**0.0**	74
tai30b	637117113	8.78	13	0.260	9	**0.0**	5
tai35a	2422002	4.95	18	2.969	15	0.92	96
tai35b	283315445	5.00	18	0.343	15	**0.0**	11
tai40a	3139370	4.53	24	3.063	24	0.79	141
tai40b	637250948	7.30	24	0.280	24	**0.0**	11
tai50a	4938796	4.51	37	3.487	50	1.34	258
tai50b	458821517	5.60	37	0.291	50	**0.0**	163
tai60a	7205962	4.54	54	3.686	88	1.27	433
tai60b	608215054	5.12	54	0.313	90	**0.0**	133
tai80a	13499184	4.35	95	2.996	220	1.53	1067
tai80b	818415043	6.63	95	1.108	225	**0.0**	1033
sko42	15812	-	-	0.654	25	**0.0**	87.06
sko49	23386	-	-	0.661	45	0.05	245.17
sko56	34458	-	-	0.729	69	0.02	356.25
sko64	48498	-	-	0.504	105	0.00	521.09
sko72	66256	-	-	0.702	153	0.11	820.53
sko81	90998	-	-	0.493	222	0.09	1090.11
sko90	115534	-	-	0.591	307	0.04	1469.63

In the following three tables of experimental results, the first column 'Instance' shows the name of the instance, while the column 'Bkv' shows the optimal solution value taken from the QAPLIB [8]. In Table II, the column 'PDbest(%)' gives the percentage deviation of the best solution value over the optimal solution value of 10 runs, and the column 'time' denotes the average run time. In Table III, the remaining columns 'mean' and 'Stdev' are the mean and standard deviation of solution cost (expressed as Average Percentage Deviation (APD), which is the percentage by which the cost exceeds the Bkv). The last Table IV which gives statistic results of testing RKCS on some QAP instances contains nine columns. The column 'best' shows the value of the best found solution. The column 'average' gives the average solution value of the 10 independent runs, the column 'worst' shows the value of the worst solution found by RKCS. The column 'SD' denotes the standard deviation. The column 'PDav (%)' denotes the percentage deviation of the average solution value over the optimal solution value of 10 runs. Bold values indicate that the solutions found have the same length as the Bkv. The percentage deviation of a solution to the best known solution (or optimal solution if known) is given by the Equation 9.

$$PDsolution(\%) = \frac{solution\ value - best\ known\ solution\ value}{best\ known\ solution\ value} \times 100 \quad (9)$$

The three tables show that RKCS can be a useful tool to switch from continuous search space to combinatorial one. It also facilitates a better control of the balance between intensification and diversification through local and global random walks.

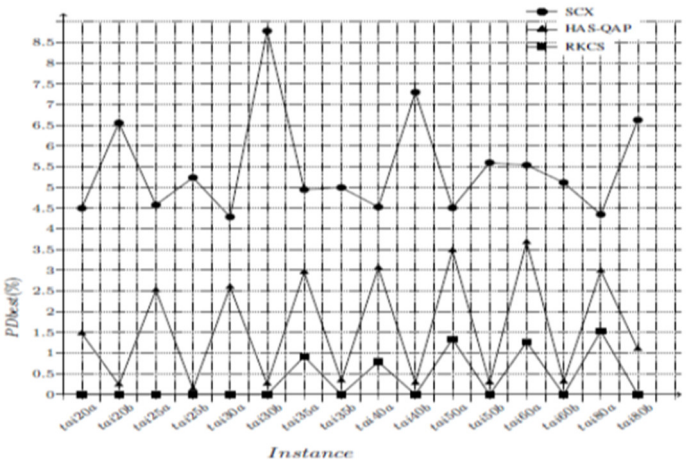


Fig. 3. PDbest(%) for 16 QAPLIB instances.

SCX performs a modified sequential constructive crossover operators in genetic algorithms, without using any robust local search technique to improve the solution quality. HAS-QAP is a hybrid ant colony system coupled with a simple local search that consists in applying a complete neighborhood examination twice with first improving strategy [15]. These two methods are compared with RKCS in Table II and Figure 3. We can observe that the difference between RKCS and the two methods is significant. This can be explained basically by a good balance between intensification and diversification, an intelligent use of Lévy flights and the reduced number of parameters.

Table III, shows the comparison of RKCS and HPSO. In this hierarchical particle swarm optimization, particles are arranged in a tree-like hierarchy, where the better-performing particles float towards the top of the hierarchy, and each particle is influenced by its immediate parent in the tree [16]. We have used in this comparison, the mean and the standard deviation. The results prove that RKCS outperforms HPSO in solving the tested QAP instanced. One of the advantages of the RKCS is the relatively independent of cuckoos in the search process for the best solution, and the use of several search strategies adopted by each category of cuckoos, without using any advanced local search method.

These results show that RKCS can be adapted easily to solve QAP. Therefore, we can say that the random-key en- coding scheme can be a very useful tool for switching from continuous to combinatorial spaces. It allows operators of the continuous space to behave freely, then projecting the changes made by these operators in the combinatorial space. It also facilitates a better control in balancing intensification and diversification through Lévy flights, which make intensified small steps in a limited region followed by a big explorative jump to a distant region. Using the real numbers, Lévy flights can easily act with the notion of distance, and can define clearly small or big steps. Then RK projects these changes in the space of QAP solutions.

TABLE III
COMPARISON OF RKCS WITH HPSO A PARTICLE SWARM OPTIMISATION BASED ALGORITHM [16].

		HPSO		*RKCS*				*HPSO*		*RKCS*	
Instance	Bkv	mean	Stdev	mean	Stdev	Instance	Bkv	mean	Stdev	mean	Stdev
bur26a	5426670	1.9	0.22	**0.0**	**0.0**	tai12a	224416	8.11	1.42	**0.0**	**0.0**
bur26b	3817852	1.96	0.41	**0.0**	**0.0**	tai12b	39464925	4.82	1.79	**0.0**	**0.0**
bur26c	5426795	2.3	0.24	**0.0**	**0.0**	tai15a	388214	7.28	0.89	0.02	0.00
bur26d	3821225	2.18	0.42	**0.0**	**0.0**	tai15b	51765268	1.05	0.15	**0.0**	**0.0**
bur26e	5386879	2.14	0.42	**0.0**	**0.0**	tai20a	703482	12.04	0.73	0.4	0.36
bur26f	3782044	2.39	0.59	**0.0**	**0.0**	tai20b	122455319	9.67	1.04	**0.0**	**0.0**
bur26g	10117172	2.07	0.22	**0.0**	**0.0**	tai25a	1167256	11.62	0.45	0.81	0.62
bur26h	7098658	2.27	0.4	**0.0**	**0.0**	tai25b	344355646	27.12	4.34	**0.0**	**0.0**
chr25a	3796	168.41	11.79	0.25	0.25	tai30a	1818146	12.25	0.41	0.89	0.35
els19	17212548	27.04	4.91	**0.0**	**0.0**	tai30b	637117113	25.18	3.28	**0.0**	**0.0**
kra30a	88900	26.85	1.09	0.16	0.16	tai35a	2422002	13.19	0.38	1.24	0.30
kra30b	91420	24.98	1.23	0.02	0.02	tai35b	283315445	27.86	2.42	**0.0**	**0.0**
wil50	48816	8.69	0.26	0.04	0.04	tai40a	3139370	13.52	0.32	1.23	0.37
nug30	6124	16.85	0.79	0.04	0.04	tai40b	637250948	35.94	2.3	**0.0**	**0.0**
nug20	2570	11.65	0.95	**0.0**	**0.0**	tai50a	4938796	13.87	0.27	1.63	0.22
sko42	15812	16.54	0.46	0.03	0.03	tai50b	458821517	34.88	1.9	0.00	0.00
sko49	23386	15.52	0.42	0.15	0.07	tai60a	7205962	13.54	0.3	1.69	0.38
sko56	34458	15.88	0.43	0.15	0.12	tai60b	608215054	36.5	1.28	**0.0**	**0.0**
sko64	48498	14.68	0.42	0.14	0.12	tai80a	13499184	12.46	0.18	1.81	0.26
sko72	66256	14.52	0.26	0.24	0.11	tai80b	818415043	34.48	0.94	0.07	0.06
sko81	90998	13.98	0.28	0.23	0.13	tai100a	21052466	11.68	0.14	1.58	0.25
sko90	115534	13.79	0.23	0.25	0.16	tai100b	1185996137	33.03	0.96	0.23	0.13

TABLE IV
RESULTS OF APPLYING RKCS ON A SET OF QAP INSTANCES.

Instance	Bkv	Best	Average	Worst	SD	PDav(%)	PDbest(%)	time(s)
bur26a	5426670	**5426670**	5426670.00	5426670	0.0	0.0	0.0	0.55
bur26b	3817852	**3817852**	3817852.00	3817852	0.0	0.0	0.0	0.76
bur26c	5426795	**5426795**	5426795.00	5426795	0.0	0.0	0.0	1.16
bur26d	3821225	**3821225**	3821225.00	3821225	0.0	0.0	0.0	1.77
bur26e	5386879	**5386879**	5386879.00	5386879	0.0	0.0	0.0	1.18
bur26f	3782044	**3782044**	3782044.00	3782044	0.0	0.0	0.0	0.60
bur26g	10117172	**10117172**	10117172.00	10117172	0.0	0.0	0.0	1.73
bur26h	7098658	**7098658**	7098658.00	7098658	0.0	0.0	0.0	0.75
chr25a	3796	**3796**	3805.75	3874	9.75	0.25	0.0	13.76
els19	17212548	**17212548**	17212548.00	17212548	0.0	0.0	0.0	0.22
kra30a	88900	**88900**	89048.75	90090	148.75	0.16	0.0	9.56
kra30b	91420	**91420**	91446.25	91490	26.25	0.02	0.0	30.05
wil50	48816	**48816**	48839.75	48886	23.75	0.04	0.0	220.50
nug30	6124	**6124**	6126.50	6128	2.5	0.04	0.0	39.90
nug20	2570	**2570**	2570.00	2570	0.0	0.0	0.0	0.67
tai12a	224416	**224416**	224416	224416	0.0	0.0	0.0	0.02
tai12b	39464925	**39464925**	39464925	39464925	0.0	0.0	0.0	0.03
tai15a	388214	**388214**	388224.8	388250	12.97	0.02	0.0	4.16
tai15b	51765268	**51765268**	51765268	51765268	0.0	0.0	0.0	0.06
tai20a	703482	**703482**	706342.0	708654	2565.75	0.40	0.0	19.94
tai20b	122455319	**122455319**	122455319	122455319	0.0	0.0	0.0	0.44
tai25a	1167256	**1167256**	1176732.8	1181894	7316.99	0.81	0.0	39.45
tai25b	344355646	**344355646**	344355646	344355646	0.0	0.0	0.0	1.13
tai30a	1818146	**1827982**	1834456.2	1843032	6474.19	0.89	0.54	74.01
tai30b	637117113	**637117113**	637117113	637117113	0.0	0.0	0.0	4.93
tai35a	2422002	2444344	2452223.4	2460532	7270.30	1.24	0.9	95.63
tai35b	283315445	**283315445**	283315445	283315445	0.0	0.0	0.0	10.52
tai40a	3139370	3164372	3178239.4	3192798	11917.91	1.23	0.79	140.86
tai40b	637250948	**637250948**	637250948	637250948	0.0	0.0	0.0	10.69
tai50a	4938796	5005070	5019579.6	5027586	11282.49	1.63	1.34	258.518
tai50b	458821517	**458821517**	458842516.4	458969332	20999.39	0.00	0.0	163.532
tai60a	7205962	7297618	7328396.4	7359568	27787.29	1.69	1.27	433.493
tai60b	608215054	**608215054**	608215054.0	608215054	0.0	0.0	0.0	133.697
tai64c	1855928	**1855928**	1855928.0	1855928	0.0	0.0	0.0	115.06
tai80a	13499184	13705886	13744806.8	13781908	35650.82	1.81	1.53	1067.45
tai80b	818415043	**818415043**	818994242.5	823187691	571643.11	0.07	0.0	1033.37
tai100a	21052466	21322258	21385568.0	21421678	53170.40	1.58	1.28	1993.95
tai100b	1185996137	1187179912	1188877434.9	1189986795	1561365.85	0.23	0.09	2221.83
sko42	15812	**15812**	15816.7	15844	4.75	0.03	0.0	87.06
sko49	23386	23398	23421.5	23450	17.80	0.15	0.05	245.17
sko56	34458	34466	34511.0	34584	42.16	0.15	0.02	356.25
sko64	48498	48502	48570.5	48686	60.57	0.14	0.00	521.09
sko72	66256	66332	66420.0	66490	79.25	0.24	0.11	820.53
sko81	90998	91088	91208.7	91380	120.75	0.23	0.09	1090.11
sko90	115534	115582	115832.2	116042	188.99	0.25	0.04	1469.63
sko100a	152002	152324	152470.2	152692	121.95	0.30	0.21	2287.07

7 Conclusion

Random Key Cuckoo Search (RKCS) is designed to be easily adapted to solve many combinatorial optimization problems such as the Travelling Salesman Problem and the Quadratic Assignment Problem. Indeed, we have shown that it is easy to move in the continuous space by using Le´vy flights and project these moves in appropriate combinatorial space, of the treated problem, via random key. It proved a good balance between intensification and diversification through Le´vy flights. In this paper, we proposed an application of RKCS to solve QAP. During the design of RKCS, we have focused on the possibility of facilitating the reuse of this metaheuristic in a direct adaptation to solve various combinatorial optimization problems.

The results are promising, but there is still room for improvement. For example, for the moment, the random key is uniform, and it may be useful to investigate if how any non- uniform decoding of random keys may affect the performance.

In addition, it will be also useful to extend the proposed approach to study larger scale benchmarks. Furthermore, the proposed RKCS may also be useful to solve other combina- torial optimization problems.

REFERENCES

[1] Ahmed, Z.: A simple genetic algorithm using sequential constructive crossover for the quadratic assignment problem. Journal of Scientific and Industrial Research (2014)

[2] Avriel, M.: Nonlinear programming: analysis and methods. Courier Dover Publications (2012)

[3] Bean, J.: Genetic algorithms and random keys for sequencing and optimization. ORSA journal on computing 6, 154–154 (1994)

[4] Blum, C., Roli, A.: Metaheuristics in combinatorial optimization: Overview and conceptual comparison. ACM Computing Surveys (CSUR) 35(3), 268–308 (2003)

[5] Brixius, N.W., Anstreicher, K.M.: The Steinberg wiring problem. SIAM (2001)

[6] Brown, C.T., Liebovitch, L.S., Glendon, R.: Le´vy flights in dobe ju/?hoansi foraging patterns. Human Ecology 35(1), 129–138 (2007)

[7] Burkard, R.E.: Quadratic assignment problems. Springer (2013)

[8] Burkard, R.E., Karisch, S.E., Rendl, F.: Qap–a quadratic assignment problem library. Journal of Global Optimization 10(4), 391–403 (1997)

[9] Burkard, R.E., Offermann, D.M.J.: Entwurf von schreibmaschinen- tastaturen mittels quadratischer zuordnungsprobleme. Zeitschrift fu¨ r Operations Research 21(4), B121–B132 (1977)

[10] Christofides, N., Benavent, E.: An exact algorithm for the quadratic assignment problem on a tree. Operations Research 37(5), 760–768 (1989)

[11] Demirel, N.C¸ ., Toksarı, M.D.: Optimization of the quadratic assignment problem using an ant colony algorithm. Applied Mathematics and Computation 183(1), 427–435 (2006)

[12] Eiselt, H.A., Laporte, G.: A combinatorial optimization problem arising in dartboard design. Journal of the Operational Research Society pp. 113–118 (1991)

[13] Elshafei, A.N.: Hospital layout as a quadratic assignment problem. Operational Research Quarterly pp. 167–179 (1977)

[14] FlNKE, G., Burkard, R., Rendl, F.: Quadratic assignment problems. Surveys in combinatorial optimization p. 61 (2011)

[15] Gambardella, L.M., Taillard, E., Dorigo, M.: Ant colonies for the quadratic assignment problem. Journal of the operational research society pp. 167–176 (1999)

[16] Helal, A.M., Abdelbar, A.M.: Incorporating domain-specific heuristics in a particle swarm optimization approach to the quadratic assignment problem. Memetic Computing 6(4), 241–254 (2014)

[17] Hochbaum, D.S.: Approximation algorithms for NP-hard problems. PWS Publishing Co. (1996)

[18] Krarup, J., Pruzan, P.M.: Computer-aided layout design. In: Mathemat- ical programming in use, pp. 75–94. Springer (1978)

[19] Layeb, A.: A novel quantum inspired cuckoo search for knapsack problems. International Journal of Bio-Inspired Computation 3(5), 297–305 (2011)

[20] Li, X., Yin, M.: A hybrid cuckoo search via le´vy flights for the permutation flow shop scheduling problem. International Journal of Production Research 51(16), 4732–4754 (2013)

[21] Mamaghani, A.S., Meybodi, M.R.: Solving the quadratic assignment problem with the modified hybrid pso algorithm. In: Application of Information and Communication Technologies (AICT), 2012 6th International Conference on, pp. 1–6. IEEE (2012)

[22] Maniezzo, V., Colorni, A.: The ant system applied to the quadratic as- signment problem. Knowledge and Data Engineering, IEEE Transactions on 11(5), 769–778 (1999)

[23] Misevicius, A.: Restart-based genetic algorithm for the quadratic assign- ment problem. In: Research and Development in Intelligent Systems XXV, pp. 91–104. Springer (2009)

[24] Nugent, C.E., Vollmann, T.E., Ruml, J.: An experimental comparison of techniques for the assignment of facilities to locations. Operations research 16(1), 150–173 (1968)

[25] Ouaarab, A., Ahiod, B., Yang, X.S.: Discrete cuckoo search algorithm for the travelling salesman problem. Neural Computing and Appli- cations pp. 1–11 (2013). DOI 10.1007/s00521-013-1402-2. URL http://dx.doi.org/10.1007/s00521-013-1402-2

[26] Ouaarab, A., Ahiod, B., Yang, X.S.: Improved and discrete cuckoo search for solving the travelling salesman problem. In: Cuckoo Search and Firefly Algorithm, pp. 63–84. Springer (2014)

[27] Ouaarab, A., Ahiod, B., Yang, X.S.: Random-key cuckoo search for the travelling salesman problem. Soft Computing pp. 1–8 (2014). DOI 10.1007/s00500-014-1322-9. URL http://dx.doi.org/10.1007/s00500-014-1322-9

[28] Shlesinger, M.F., Zaslavsky, G.M., Frisch, U.: Le´vy flights and related topics in physics. In: Levy flights and related topics in Physics, vol. 450 (1995)

[29] Skorin-Kapov, J.: Tabu search applied to the quadratic assignment problem. ORSA Journal on computing 2(1), 33–45 (1990)

[30] Taillard, E.: Robust taboo search for the quadratic assignment problem. Parallel computing 17(4), 443–455 (1991)

[31] Taillard, E.: Robust taboo search for the quadratic assignment problem. Parallel computing 17(4), 443–455 (1991)

[32] Taillard, E.D.: Comparison of iterative searches for the quadratic assign- ment problem. Location science 3(2), 87–105 (1995)

[33] Wilhelm, M.R., Ward, T.L.: Solving quadratic assignment problems by simulated annealing. IIE transactions 19(1), 107–119 (1987)

[34] Yang, X.S., Deb, S.: Cuckoo search via le´vy flights. In: Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on, pp. 210–214. IEEE (2009)

